OUTLOOK
Genome editing

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

SAVE OUR SOILS

The make-up and management of soils and their influence on the environment and human health PAGES 32, 51, 60 & 69

RESEARCH TRAINING

DOCTORATE IN DISTRESS

How to build a better PhD system

PAGE 22

PUBLIC HEALTH

EVERY BREATH YOU TAKE ...

Wearable body sensors could transform health care

PAGE 26

SUSTAINABILITY

WASTE NOT, WANT NOT

Mine water pollutants for valuable elements

PAGE 29

NATURE.COM/NATURE
3 December 2015 £10
Vol. 528, No. 7580

EDITORIALS

Make the most of PhDs

The number of people with science doctorates is rapidly increasing, but there are not enough academic jobs for them all. Graduate programmes should be reformed to meet students' needs.

It is hard to argue against the idea that a workforce should be highly educated. The media, politicians and universities all believe that a scientific background will not only benefit individuals, but also drive science, innovation and the economy. As a result, the number of people entering higher education in the sciences and engineering has been on the rise for decades. Between 1995 and 2012, the Organisation for Economic Co-operation and Development reported an overall increase in university graduation rates of 22 percentage points. In the same time frame, the PhD production rate has doubled, even though PhDs account for only a small percentage of higher-education graduations.

Getting a science PhD can be a very fulfilling experience, in which students spend a few years enjoying the rigour and freedom of academic research. Many pursue a PhD because they love the science, to satisfy their curiosity about the world or to contribute to a growing body of knowledge. All hope to emerge with the skills to pursue career goals, within or outside academia.

They have a good chance of doing so, too. Science PhD holders experience very low unemployment rates — just 2.1% of people with doctorates in science, engineering or health in the United States were unemployed in 2013, according to the National Science Foundation (NSF) Survey of Doctorate Recipients. The overall national unemployment rate for people aged 25 or older was 6.3%.

But the chances of getting a faculty job in academia — the career dream of many — are slim. Of the employed doctorate holders in the NSF survey, just over 50% were working outside academia across a variety of sectors, including industry, federal government and non-profit organizations. Many young researchers feel that their graduate training does not adequately prepare them for these different careers. Nor do they feel that they are being properly informed of their future prospects or the realities of the training. Many principal investigators, universities, funding bodies and governments are keen to keep pushing the message that a science PhD is good, and that there are plenty of jobs in academia. Job markets are not fixed, and can change substantially between the start and end of a lengthy PhD.

The opportunity cost of a PhD can also be substantial for young people. Not all have the luxury of being able to spend several years on a PhD with low pay and no clear destination — they can't afford it, might miss out on other opportunities, or prefer to pursue deep training in another sphere that is more appropriate for their skills and chosen careers.

If we accept that there are positives to having lots of PhD holders, then we need to work out how the system should change to support them all. As a News Feature on page 22 explores, various suggestions are bouncing around. One is to revamp the PhD so that it combines research with development of workplace skills. Several institutions, such as the University of California, San Francisco, run courses that offer graduate students training in management, communication and entrepreneurship.

Students could also skip the PhD entirely. Many who are contemplating a doctorate but aren't sure of its value to their future could instead experience postgraduate research through a master's degree.

A more controversial idea — around since the 1970s — is to cut the total number of graduate students entering the system. This has met with stiff resistance from faculty members, university funding bodies and governments.

"If we accept that there are positives to having lots of PhD holders, then we need to work out how the system should change to support them."

The biggest problem for early-career researchers seems to be a lack of data on the career trajectories and opportunities available to them. Although some information-gathering efforts exist, none is substantial enough to provide the detail needed for students to make informed decisions about their futures.

To create a happy, sustainable PhD population, collaborative efforts between students, academics, industry and government leaders are needed. A science PhD is no longer an apprenticeship in science for academia, but an apprenticeship in scientific thinking that is beneficial for all walks of life. There are already some grass-roots campaigns in this direction, but they are not enough. The welfare and future of the economy and science rest on the shoulders of young, highly educated workers. Policymakers need to start putting the graduates' needs first. ■

Root causes

Research has a part to play in identifying the factors that breed terrorism.

How can research best contribute to understanding and fighting terrorism, and to improving counterterrorism policies? A common focus is a narrow concept of radicalization that explores why individuals turn to extremism.

Since the 11 September terrorist attacks in the United States in 2001 — and the deadly bombings in Madrid in 2004 and London in 2005 — an entire industry of government-funded consultants and researchers has grown up around this idea. But many researchers find such emphasis problematic; they argue, for example, that it can distract from the need for a broader understanding of the roots of terrorism. They also fear that counterterrorism policies based on it may be ineffective, and risk being counterproductive.

Research to understand why and how people, such as the young people who carried out the attacks in Paris on 13 November, become

radicalized is crucial, and as a News story on page 20 describes, such research has provided some important insights. But there is no typical profile of those who turn to violent extremism, and the causes are highly diverse. Radicalization has become a central plank in national counterterrorism policies, with efforts made to identify individuals and groups showing signs of radicalization or vulnerability, and to de-radicalize them.

Under Britain's 'Prevent' programme, the government earlier this year made it compulsory for staff in schools, universities, councils, prisons and other bodies to monitor or refer such individuals to the authorities. Yet it's clear that only a tiny minority of the vast numbers of people flagged by counter-radicalization efforts risk turning to terrorism — and spotting which ones is extremely difficult.

Some researchers argue that such policies are justified on the grounds that the immediate terrorist threat to democracies is so great that everything must be done in the short term to stop people from becoming radicalized and to spot violent extremists, while also addressing the broader causes and dynamics of terrorism. But other researchers question the effectiveness of such policies, and argue that the focus should be more on community policing and on reinforcing intelligence to identify the recruiters and ringleaders of terrorist networks. Social profiling, they add, comes at a potentially high cost. It risks stigmatizing further Muslims and those of immigrant origin, and inadvertently legitimizing the anti-Islam, and often racist, rhetoric of extreme-right-wing parties. The resulting social division risks making matters worse and increasing the pool of potential terror recruits.

As Nadia Fadil, who specializes in Islam in Europe at the University of Leuven in Belgium, points out, policies based on targeting Muslim populations also risk harming existing community-based prevention methods. In Belgium, youth workers, teachers and other officials who were previously considered trustworthy bridge-builders are increasingly distrusted in the communities in which they work because they are now perceived as state spies.

Shortcomings of radicalization research itself were highlighted in a 2013 review of the research literature by Alex Schmid, a director of the Terrorism Research Initiative, an international consortium of researchers and research centres. It concluded that the search for the causes of radicalization of young people has produced "inconclusive results", and that counter-radicalization and de-radicalization programmes lack rigorous evaluation.

"There is no typical profile of those who turn to violent extremism, and the causes are highly diverse."

Most worryingly, the review highlighted blind spots in radicalization research. Much, it concluded, is "one-sided", in that it looks only at the radicalization of Islamist, non-state actors, and ignores the fact that radicalization of Western governments can also occur, combining in a vicious circle that can fuel strife and terrorism.

That is an unpopular view, and is often considered by politicians and the media to be making excuses for terrorism. But it must be taken into account to develop more effective policies and to identify those that are ineffective or even harmful.

Research can do its bit, by bringing an evidence-based, neutral and broader perspective that can enlighten counterterrorism, social, educational and other policies. That need is now greater than ever. ■

Take more risks

Scientific innovation is being smothered by a culture of conformity.

Suppose you are devising a technique to transfer proteins from a gel to a plastic substrate for easier analysis. Useful, maybe — but will you gain kudos for it? A notable finding of last year's survey of the 100 most cited papers on the Web of Science (see *Nature* **514,** 550; 2014) was how many of them reported such apparently mundane methodological research (this protein-transfer method came in at number six).

Not all prosaic work reaches such bibliometric heights, but that does not deny its value. Overcoming the hurdles of nanoparticle drug delivery, for example, requires the painstaking characterization of pathways and rates of breakdown and loss in the body — work that is probably unpublishable, let alone unglamorous. One can cite comparable demands for detail to get just about any bright idea to work in practice — but it's usually the initial idea, not the hard grind, that garners the praise. The incentives for such boring but essential collection of fine-grained data to solve a specific problem are vanishing in a publish-or-perish culture.

Meanwhile, a recent analysis of discovery and innovation in biomedicine, using the molecules studied as value markers, finds that the choice of research problems is becoming more conservative and risk-averse (A. Rzhetsky *et al. Proc. Natl Acad. Sci. USA* **112,** 14569–14574; 2015). One might quibble with the scope of the study, but its general conclusions — that current norms discourage risk and therefore slow down scientific advance, and that the problem is worsening — ring true.

Attempts to hit the publishable 'sweet spot' by avoiding both the prosaic and the risky are likely to reduce the efficiency of scientific discovery. But a fashionably despairing cry of 'Science is broken!' is not the way forward. The wider virtue of Rzhetsky *et al.*'s study is that it floats the notion of tuning practices and institutions to accelerate the process of scientific discovery. The researchers conclude, for example, that publication of experimental failures would assist this goal by avoiding wasteful repetition. Journals chasing impact factors might not welcome that, but they are no longer the sole repositories of scientific findings. Rzhetsky *et al.* also suggest some shifts in institutional structures that might help promote riskier, but potentially more groundbreaking, research — for example, spreading both risk and credit among teams or organizations.

The danger is that efforts to streamline discovery simply become codified into another set of guidelines and procedures, creating yet more hoops for grant applicants to jump through.

A better first step would be to recognize the message that research on complex systems has emphasized: efficiencies are much more likely to come from the bottom up. The aim is to design systems with basic rules of engagement for participating agents that best enable an optimal state to emerge. Such principles typically confer adaptability, diversity and robustness. There could be a wider mix of grant sources and sizes, say, less rigid disciplinary boundaries, and wider acceptance that citation records are not the only measure of worth.

But perhaps more than anything, the current narrowing of objectives, opportunities and strategies in science reflects an erosion of trust. Obsessive focus on 'impact' and regular scrutiny of bibliometric data betray a lack of trust that would have sunk many discoveries and discoverers of the past. Bibliometrics might sometimes be hard to avoid as a first-pass filter for appointments (see *Nature* **527,** 279; 2015), but a steady stream of publications is not the only, or even the best, measure of potential.

Attempts to tackle these widely acknowledged problems are typically little more than a timid rearranging of deckchairs. Partly that's because they are seen as someone else's problem: the culprits are never the complainants, but the referees, grant agencies and tenure committees who oppress them. Yet oddly enough, these obstructive folk are, almost without exception, scientists too (or at least, they once were). Inefficiencies can exact a huge price. It is time to oil the gears. ■

➲ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

WORLD VIEW *A personal take on events*


ANNE KOERBER/LSHTM

The world must accept that the HPV vaccine is safe

But the science alone will not be enough to build public and political confidence, says **Heidi Larson**.

Every year brings 528,000 new cases of cervical cancer and 266,000 deaths, linked to human papillomavirus (HPV). We have a highly effective HPV vaccine, but suspicion stands in the way of its adoption in many countries. How can we dispel this mistrust?

On 20 November, a report from the European Medicines Agency (EMA) confirmed the vaccine's safety. The agency had been asked by Denmark to reinvestigate after symptoms of dizziness, fainting, aches and pains were reported in adolescent girls and suspicion fell on the vaccine. It is not the only country to report such events.

The good news is that public concern about these reactions is being heard and has prompted further investigation. The EMA report is one of many to confirm the safety of the vaccine and conclude that there is no need to change vaccination policies.

The not-so-good news is that not everyone believes them.

Evidence suggests that the events were 'psychogenic illnesses', psychological reactions that can spread fast, especially when girls are vaccinated in groups at school and witness each other's reactions. A growing collection of YouTube clips is also fuelling anxieties.

My research group studies situations in which public, provider or political trust in vaccines has been broken. We have heard many testimonies of the anxiety that politicians and decision-makers face when pressured about suspected vaccine reactions while also hearing that scientific evidence exonerates the vaccines. We have learned the importance of monitoring public sentiment, responding promptly to concerns and engaging and listening to the public early on when vaccines are being introduced.

In some nations, politicians side with the science. In others, they bend to minority opinions. Japan reacted ambiguously to reports of HPV vaccine side effects: it withdrew 'proactive' recommendation of the vaccine while it investigated, but continued to provide the vaccine for those who demanded it. The investigations found no clear causal link to the vaccine, but the recommendation remains suspended.

In another case, in 2010, we investigated the suspension of HPV vaccine demonstration projects in two Indian states. Vaccination acceptance was high in the projects; the pressure had come from an activist women's group far away in New Delhi. When the group's demands for public dialogue about the safety, efficacy and cost-effectiveness of the initiative were not answered, it found, and widely reported, seven deaths among girls who had participated.

These deaths were judged unrelated to the vaccine, but the projects never resumed. Nearly five years later, millions of women are missing out on the chance to prevent cervical cancer. One-quarter of global cervical-cancer deaths are in India.

NATURE.COM
Discuss this article online at:
go.nature.com/gznxgg

Some governments stand by the science even when faced with public panic. Last year, 600 girls in a Colombian municipality reported symptoms after HPV vaccination. Faced with local anxieties and some anger, the Colombian government expressed empathy, and the vaccination programme continues. England reached 87% full-dose coverage in 2014, having averted a potential public-confidence crisis in 2009, when a 14-year-old girl died after being vaccinated. Health officials expressed concern, promptly investigated the girl's death and found it unrelated to the vaccine.

Psychogenic reactions are not unique to HPV vaccination. During the 2009 H1N1 influenza pandemic, there were 23 episodes of mass psychogenic illness in Taiwan's school flu-vaccination programme. In Iran, people panicked after 10 girls in a class of 26 experienced psychogenic reactions after tetanus shots.

THERE ARE BROAD LESSONS HERE, BUT THE HPV VACCINE CARRIES UNIQUE CHALLENGES.

I learned about the Iran situation while working with UNICEF just over a decade ago, when I was asked to help plan a nationwide measles campaign — and, specifically, to design ways to pre-empt the type of panic provoked by the tetanus vaccine reactions. The measles campaign was a success, but it took considerable advance work that included gathering local input into communication materials and outreach early in their preparation; engaging young people (the campaign was targeting everyone under 25 years old); and working with schools, local leaders and the media.

The HPV vaccine carries unique challenges. Because the first thing it prevents is sexual transmission of HPV, use of the vaccine evokes moral judgements around sexual behaviour.

The United States is struggling to get HPV vaccination coverage above 40%. Some parents are anxious that the vaccine will make their daughters more promiscuous, even though multiple studies have found no such effect. Other reports cite 'embarrassment' in some cultures about accepting the vaccine.

The HPV vaccine touches nerves, and acceptance needs strategies that vary between cultural and political settings. Despite the challenges, more than 80 million girls and women around the world have received the vaccination.

We should not underestimate the potential for progress to be disrupted by the mass spread of vaccine reactions and concerns, the amplification that can follow through social media and the vulnerability of political processes, which sometimes find themselves paralysed between public and scientific opinion. ■

Heidi Larson *heads the Vaccine Confidence Project at the London School of Hygiene & Tropical Medicine, where she is a senior lecturer in the Department of Infectious Disease Epidemiology.*
e-mail: heidi.larson@lshtm.ac.uk

RESEARCH HIGHLIGHTS

Selections from the scientific literature

ASTROPHYSICS

Supernova glow shows stellar twin

The force of an exploding star may have ripped material off an orbiting companion star, leaving behind a signature glow.

Astronomers first spotted the massive explosion of supernova iPTF13ehe in 2013. Two years later, they noticed an afterglow coming from clouds of hydrogen nearby. Takashi Moriya at the University of Bonn in Germany and his colleagues argue that this unexpected light came from material that was torn off another star during the violent original outburst. The energy from the blast could have blown away part of a tightly orbiting companion star, stripping off a mass of hydrogen that could weigh nearly as much as the Sun.

Careful scrutiny of hydrogen emissions from other especially bright supernovae could determine whether this radiance stems from companion stars or matter that is already present in the surrounding interstellar space.

Astron. Astrophys. **584, L5 (2015)**

RICHARD STUPART/CC-BY-2.0

ENVIRONMENTAL SCIENCES

Ecological toll of African infrastructure

Huge development projects such as roads and railways that are planned or under construction in Africa threaten swathes of its ecosystems.

William Laurance and his team at James Cook University in Cairns, Australia, mapped 33 'development corridors' that are being upgraded or planned, plus their human populations and surrounding lands. They found that these corridors would stretch 53,000 kilometres and cut through 408 protected areas, 29 of which would be cut by two or more corridors.

Corridors are often justified on the basis of their benefits to agricultural production, but the team found just five that would have both low environmental impact and large agricultural benefit. Six would degrade areas with high conservation value and bring low agricultural benefits, and the rest would bring only "marginal" returns. Many of the developments would cause serious and irreversible damage.

Curr. Biol. **http://doi.org/9kg (2015)**

NEUROSCIENCE

Alzheimer's role of breast-cancer gene

The DNA-repair protein BRCA1 is known to increase the risk of breast and ovarian cancer when it is mutated. But the normal protein might also have a central role in Alzheimer's disease.

Elsa Suberbielle and Lennart Mucke at the Gladstone Institute of Neurological Disease in San Francisco, California, and their colleagues lowered BRCA1 protein levels in mouse brains by blocking the *BRCA1* gene using a small piece of RNA. They found that some neurons shrank in size and that the animals had a reduced ability to learn and remember their way around a maze. The researchers also showed that BRCA1 levels were depleted in the post-mortem brains of people with Alzheimer's. By looking at this process in cultured neurons and in mice, the authors suggest that BRCA1 is degraded when amyloid-β proteins accumulate in the brain in Alzheimer's disease.

Nature Commun. **http://doi.org/9kk (2015)**

GEOPHYSICS

Earth's magma in a spin

The rapid spin of the early Earth could have influenced the way that the planet solidified.

Some 4.5 billion years ago, Earth was extremely hot, covered by a molten magma ocean, and completed a full rotation in a few hours. Christian Maas and Ulrich Hansen of the University of Münster in Germany calculate that the fast rotation could have influenced how crystals settled from the magma ocean and shaped Earth's interior. They used a three-dimensional model of the formation of silicate crystals in magma, and found that a fast rotation rate created a crystal layer that settled deeper beneath the poles than under the equator.

This could have played a key part in how Earth's mantle layer eventually solidified out of the magma ocean, say the authors.

J. Geophys. Res. Solid Earth **http://dx.doi.org/10.1002/2015JB0121053 (2015)**

NATURE PRODUCTION/NATUREPL.COM

GENOMICS

Genome shows gecko evolution

The first genome of a gecko species hints at the basis of its ability to regrow tails and climb walls.

More than 1,400 species of gecko inhabit temperate areas across the world. A team led by Huanming Yang at BGI in Shenzhen and Xiaosong Gu at Nantong University, both in China, sequenced the genome of Schlegel's Japanese gecko (*Gekko japonicus*; **pictured**) and identified more than 22,000 genes. Comparisons with other reptile and vertebrate genomes show that geckos diverged from other lizards around 200 million years ago, after the split of two supercontinents.

The gecko genome harbours dozens of copies of β-keratin genes — expressed in hair-like growths called setae that help the animal to cling to vertical surfaces. Expression of two genes that make the hormone prostaglandin increased in geckos after their tails had been amputated, suggesting a role for this hormone in regeneration.

Nature Commun. **6, 10033 (2015)**

ATMOSPHERIC SCIENCE

Ozone destruction in a future climate

The potency of one of the major ozone-destroying gases could double because of future climate change.

Nitrous oxide (N_2O) leads to ozone destruction through various chemical reactions in the stratosphere, and is the main ozone-destroying gas released by human activity. Laura Revell at the Swiss Federal Institute of Technology in Zurich and her colleagues analysed the ozone-depletion potential of this gas using different scenarios of future climate change. The models showed that ozone destruction involving N_2O is made less efficient by the higher concentrations of carbon dioxide and methane that are expected in the atmosphere by 2100. However, the team found that other changes in atmospheric chemistry, temperature and air circulation by 2100 could still increase the ozone depletion potential of N_2O by as much as two-fold relative to 2000.

Geophys. Res. Lett. **http://doi.org/9h2 (2015)**

ECOLOGY

Africa's herbivores mapped

Researchers have constructed a map of Sub-Saharan Africa showing the types of plant-eating animals that grazed it some 1,000 years ago.

Gareth Hempson, of the University of Cape Town in South Africa, and his colleagues used factors such as species distribution, rainfall and vegetation patterns to model the likely biomass of 92 large herbivores across Sub-Saharan Africa around 1,000 years ago. They divided the region into areas each measuring around 12,000 square kilometres, and grouped areas that had similar biomass and animal types into four herbivore regimes. They named these 'herbivomes' after the forest duiker, the arid gazelle and the bulk feeder, with the fourth regime containing a high variety and abundance of larger species. The analysis should assist research on the loss of large plant-eaters and improve understanding of African ecology, the team says.

Science **350, 1056–1061 (2015)**

ECOLOGY

Pollination is more than bees

Other creatures visit more flowers than bees do, and may be almost as important in pollinating crops.

Romina Rader at the University of New England in Armidale, Australia, and her colleagues analysed data from 39 field studies of pollination by honey bees, other bees and other insects, including flies, beetles, moths and ants. They found that other insects carried out 25–50% of all visits to crop flowers. Although these 'non-bees' were less effective at pollinating on each visit, their increased visits made them roughly as effective as bees.

Crops such as coffee and grapefruit were almost exclusively pollinated by bees, whereas crops such as custard apples and mangoes relied almost totally on other insects. Non-bees were also found to be less affected by changes to natural habitats, so the authors suggest that these insects might provide a more robust pollination service than bees do.

Proc. Natl Acad. Sci. USA **http://dx.doi.org/10.1073/pnas.1517092112 (2015)**

SOCIAL SELECTION *Popular topics on social media*

Credit and co-authors cause chatter

Questions of paper authorship have been plaguing scientists on social media: who should come first? And who deserves to be listed at all? When it comes to papers with numerous authors, the publishing process can get messy. For instance, when Dorothy Bishop, a psychologist at the University of Oxford, UK, found herself trying to review a paper blemished with mistakes, she tweeted: "When a manuscript with 20+ authors has grammatical errors, typos and/or no page numbers, you wonder how many authors actually read it." Others took a less dark view. Deirdre Toher, a statistician at the University of the West of England in Bristol, UK, tweeted that the logistics of implementing changes from multiple researchers may have led to the mistakes, adding "with that many authors it also means that people assume that the basics are 'someone else's responsibility'."

NATURE.COM
For more on popular papers:
go.nature.com/noociy

ZOOLOGY

Pigeon leaders fly faster

Birds that lead a social group learn faster than their followers, although the leaders might not start out as the best decision-makers.

Benjamin Pettit and Dora Biro at the University of Oxford, UK, and their colleagues tracked the behaviour of 40 homing pigeons (**pictured**) as the birds navigated various routes, both individually and as a flock. They found that birds that later assumed leadership of flocks had been the fastest fliers on previous solo flights, but had not necessarily navigated the shortest and most energy-efficient routes. On later solo flights, leaders learned to navigate along direct routes more quickly than followers did. The team suggests that, among pigeons at least, leadership is based on pre-existing individual differences rather than on social preferences or optimal group decision-making.

Curr. Biol. **http://doi.org/9kb (2015)**

ZSUZSA ÁKOS

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

EVENTS

Anthrax vaccine

An anthrax vaccine has become the first to be approved by the US Food and Drug Administration (FDA) under the 'Animal Rule', which allows approval on the basis of animal tests when studies in humans are not ethical or possible. The FDA announced on 23 November that the vaccine, called BioThrax, can be used after exposure to *Bacillus anthracis*, the bacterium that causes anthrax. BioThrax was initially approved in 1970 to prevent anthrax before exposure to the bacterium. The vaccine is made by Emergent BioDefense Operations Lansing in Michigan.

Retraction data

A searchable database should soon allow systematic identification of retracted publications. Posts and article identifiers from the blog Retraction Watch will be incorporated into a web application maintained by Center for Open Science in Charlottesville, Virginia, that already tracks research activities such as posting preprints or depositing data sets. The resource will initially have about 5,000 entries, and was announced by both organizations on 24 November.

LHC heavy metal

After spending five months colliding protons following a major upgrade this year, the Large Hadron Collider (LHC) near Geneva, Switzerland, began a one-month run of experiments with heavy ions on 25 November. All main detectors at the accelerator — including ALICE, which was designed for this purpose — are now studying the state of matter known as quark–gluon plasma, which can arise when two nuclei of lead-208 collide. In these collisions, the nuclei carry a record-breaking energy of more than 1 petaelectronvolt.

BLUE ORIGIN/VIA ZUMA WIRE/REX SHUTTERSTOCK

Blue Origin gets to space and back

Commercial spaceflight company Blue Origin — the brainchild of Jeff Bezos, head of online retail giant Amazon — completed a test of its reusable rocket on 23 November. The autonomous vehicle was successfully landed after it propelled a capsule to a height of more than 100 kilometres, which is classed as being in space. The flight comes just seven months after one of the company's rockets was destroyed during a similar test. Blue Origin has not yet completed a crewed flight; the capsule is designed to carry up to six passengers into space.

RESEARCH

Emissions stall

Humanity's greenhouse-gas output increased by just 0.5% in 2014, despite significant global economic growth, according to figures released on 25 November. Carbon emissions rose by 3–4% per year in the first decade of the twenty-first century, but that growth has slowed dramatically over the past 3 years, report the Netherlands Environmental Assessment Agency and the European Commission's Joint Research Centre. The biggest factor is China, where slower economic growth and a shift towards cleaner energy sources and less energy-intensive manufacturing have reduced the energy intensity of the economy. See go.nature.com/kphlae for more.

Deforestation rises

The rate of legal deforestation in the Amazon rainforest has risen over the past year, Brazilian environment minister Izabella Teixeira announced on 26 November. Satellite images show that 5,831 square kilometres of forest were lost to activities such as livestock farming and agriculture in the year up to July 2015, a 16% increase on the previous year. The increases were largest in the states of Rondônia, Mato Grosso and Amazonas. Of these, Mato Grosso had the biggest area of forest loss, at 1,508 square kilometres. Efforts by the Brazilian federal government have generally been bringing down rates of deforestation, and the current rate is around one-fifth of that in 2004.

FUNDING

Energy partnership

A group of 28 investors from 10 countries has launched a multibillion-dollar clean-energy research partnership. The Breakthrough Energy Coalition, spearheaded by Microsoft founder Bill Gates, and including Virgin

founder Richard Branson and Amazon boss Jeff Bezos, was announced on 30 November, on the opening day of the international climate-change negotiations in Paris. The private partnership aims to support early-stage research into low-carbon technologies for future energy supply. It will complement energy-research efforts announced by US President Barack Obama and French President François Hollande on the same day, dubbed 'Mission Innovation'. See go.nature.com/wzigmx for more.

STAFF/REUTERS/CORBIS

POLICY

Carbon plan canned

On 25 November, the UK government scrapped a £1-billion (US$1.5-billion) competition to build a demonstration carbon capture and storage plant. Funding for the project — intended to demonstrate that carbon dioxide can be filtered out of power-plant exhaust gases on a commercial scale — has been on the table since 2012, but was removed from government plans in the latest five-year spending review.

Rhino-horn ban

A South African court has lifted a ban on the domestic trade in rhino horn (**pictured**) after two game farmers claimed that it infringed their

right to trade in a renewable substance. On 26 November, the judge ruled that the ban, introduced in 2009, had not undergone proper public consultation. He added that since 2008 the number of South African rhinos poached for their horns has increased from less than 100 per year to around 1,200. Conservation group Save the Rhino asked how a national ban could fuel poaching, which mainly serves overseas markets, given that the international trade is illegal. The South African government is to appeal the ruling; the law will stay in place until the appeal has been heard.

Open-access policy

The Netherlands Organisation for Scientific Research (NWO) is tightening its open-access policy to demand that research results become universally available as soon as authors publish them. NWO-funded researchers were previously obliged either to publish in an open-access journal or to submit a version of their work to a public database 'as soon as possible' after publishing in a pay-to-read journal. From 1 December, new grant conditions require Dutch researchers to make work immediately accessible. To avoid conflicting with journals that enforce embargo periods, such as *Nature*, researchers can submit pre-peer-review versions to a database.

FACILITIES

Animal clones

A huge animal-cloning centre in Tianjin, China, will open early in 2016. Launched with 200 million yuan (US$31.3 million) from Sinica, a subsidiary of BoyaLife in Wuxi, the Tianjin International Joint Academy of Biomedicine, Peking University in Beijing and Sooam Biotech in Seoul, the centre will clone cattle, dogs and racehorses. BoyaLife says that the aim is to produce one million cloned cow embryos annually to help Chinese farmers to meet demand for beef.

Italian expo

The Italian government enacted a decree on 25 November that allocates €80 million (US$85 million) to launch a major research centre to focus on big-data exploitation in health and nutrition, as well as nanotechnologies. Called Human Technopole, the centre will take over part of the site used for the 2015 international exhibition called Milan Expo. It will continue the theme of the exhibition — 'feeding the planet, energy for life'. Human Technopole will be led by the Genoa-based Italian Institute of Technology and will eventually employ more than 1,000 researchers.

PEOPLE

Maurice Strong

Maurice Strong, the founding head of the United Nations Environment Programme (UNEP) and a leading figure in climate-change politics, has died aged 86. He was a major figure in organizing the 1992 Rio Earth Summit and creating the UN Framework Convention on Climate Change. Strong is regarded as one of the most important people in the history of the environmental and sustainability movements. In a statement released by UNEP on 28 November, Achim Steiner, the current head of the agency, called him a visionary and a pioneer of global sustainable development.

COMING UP

3–4 DECEMBER
The first International Workshop on Metamaterials-by-design takes place in Paris.
go.nature.com/hkzg9s

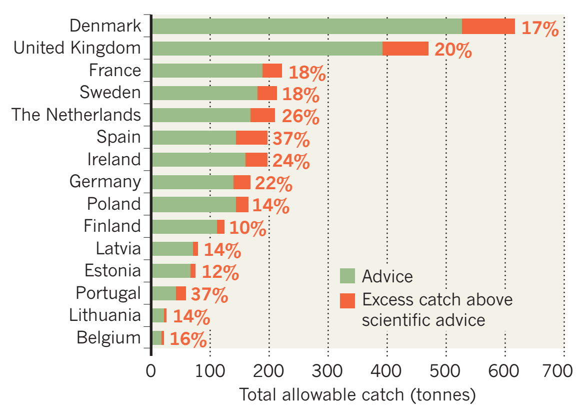
8–9 DECEMBER
The Royal Society of Medicine and the Nutrition Society in London jointly host a meeting that will look at the role of sleep in obesity and nutrition.
go.nature.com/xigwne

TREND WATCH

European politicians often allow more fish to be taken from the seas than is recommended by scientists. Yet this excess varies by country, according to a study (G. Carpenter *et al. Mar. Policy* **64,** 9–15; 2016). In 2001, the total catch permitted in the European Union averaged 33% more than that advised by the International Council for the Exploration of the Sea. In 2015, this fell to 7% above the advised level. EU politicians negotiate catch limits in secret, but more transparency is needed, say the authors.

OVERFISHING IN EUROPE

On average, Denmark and the United Kingdom received the highest catch allocations in excess of scientific advice between 2001 and 2015.



SOURCE: G. CARPENTER *ET AL. MAR. POLICY* **64**, 9–15 (2016)

NATURE.COM
For daily news updates see:
www.nature.com/news

FRANCO ORIGLIA/GETTY

The Italian Senate has approved a budget amendment that would award €3 million to a stem-cell trial.

POLITICS

Italian scientists slam trial selection

Senate assigns a stem-cell trial €3 million — but researchers call instead for an open competition.

BY ALISON ABBOTT

Italian politicians have kindled the wrath of some biomedical scientists by hand-picking a stem-cell clinical trial for funding.

The €3-million (US$3.2-million) pot for the trial should be allocated through an open competition based on scientific merit rather than in an amendment to the country's 2016 budget bill, say the researchers. They are appealing to Italy's Parliament to change the amendment — which the Senate approved on 20 November — before it passes into law.

To many Italian scientists, the idea that certain projects receive funding at the whim of politicians feels depressingly familiar. In 2013, a government decree earmarked this money for a stem-cell clinical trial run by a specific clinic. That earlier allocation was to the controversial Stamina Foundation in Brescia, and was later abandoned after a fraught, year-long campaign by scientists who convinced the health ministry that the trial was based on bad science and illegal practices. In the objections to the amendment to the 2016 budget bill, there is no suggestion of scientific wrongdoing, but researchers take issue with the way that the trial selection was made.

"Italian politicians can take a liking to a project and then finance it directly," says Marino Zerial, a director at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany. He is one of 29 researchers who signed a letter published in the newspaper *La Stampa* on 26 November calling for the budget-law amendment to be changed. "Where do you see this in any other country?"

After the Stamina trial was abandoned, many Italian scientists assumed that politicians would never again earmark specific research projects for public funding. The Constitutional Court, reflecting on the Stamina debacle, stated that the selection of clinical trials to receive public funding should not be at the "pure political discretion of the legislator".

The amendment to the 2016 budget law does not name a specific trial: it states that the money should be given for a "phase II clinical trial based on the transplantation of human neural stem cells in patients affected with Amyotrophic Lateral Sclerosis" (ALS), a deadly neurodegenerative condition also known as motor neuron disease. But of the 11 stem-cell protocols that have been approved for early-phase clinical trials in Italy, only one meets these criteria — so the outcome amounts to a selection in practice, says Giuseppe Remuzzi, director of the Mario Negri Institute for Pharmacological Research in Bergamo, one of the letter's signatories.

The protocol comes from the lab of stem-cell researcher Angelo Vescovi, scientific director of the Casa Sollievo della Sofferenza research hospital in the southern Italian province of Foggia. It involves transplanting neurons derived from the brains of miscarried fetuses into the spinal cords of people with ALS.

Senator Giorgio Santini, who proposed the amendment, told *Nature* that the way it had been selected and added to the bill was procedurally correct and that, if signed into law, the amendment would help sick people. He added that, in principle, any scientist who met the criteria could apply for the money, but that Vescovi's trial is the only one that is ready to be implemented.

Vescovi declined to comment on the fact that his work was proposed for funding, ▶

but noted that Santini was present at a public meeting in Rome on 29 September, where Vescovi presented the final results of his phase I ALS trial.

Unlike the criticisms levelled against the Stamina trial, the scientists who wrote the letter complaining about the amendment told *Nature* that they are not commenting on the quality of Vescovi's research. They point out that a phase I trial in six patients that assessed the safety of the therapy was carried out with appropriate oversight of the health authorities (L. Mazzini *et al. J. Transl. Med.* **13,** 17; 2015).

Their concerns lie with how the trial was selected, which is particularly painful because Italy's funds for research are among the lowest in Europe, and the 2016 budget foresees no increase for public institutions. "Scarce resources should be allocated to the most absolute transparency rules", they write. Remuzzi says that the criteria were so narrow that scientists like himself could not apply. "We ourselves have three stem-cell trials related to organ transplantation," he says. "We should have had the opportunity to apply."

"We should have had the opportunity to apply."

Patient groups have also objected to the way the funding was allocated, saying that therapies for other illnesses should have been allowed to compete. The Italian Multiple Sclerosis Foundation in Genoa has sent politicians evidence of other ready-to-go Italian stem-cell trials for various neurodegenerative diseases. And the Italian Association for Huntington's Chorea in Milan complained to the Chamber of Deputies — Italy's second parliamentary house, which will vote on the budget bill next — that the amendment appeared without revealing the scientific criteria used to select one clinical trial from among other candidates.

Scientists hope that their outcry will encourage the Chamber of Deputies to drop or modify the amendment in the next week or so. Time is tight: the 2016 budget law is linked to a confidence vote in the government, which means that Parliament is under pressure to approve it before the end of the year. ■

Crews lower huge drills down from specially equipped ships to penetrate the sea floor.

GEOLOGY

Drill ship targets Earth's mantle

Indian Ocean expedition resumes quest to bore right through the planet's crust.

BY ALEXANDRA WITZE

Jules Verne would have dug this plan: drill into the sea floor, through kilometres of the planet's rocky crust to penetrate the denser underlying mantle. It is one of geology's classic quests, conceived almost 60 years ago, at the peak of the plate-tectonics revolution. Since then, many have attempted it and failed. But an expedition starting this month is taking up the challenge once again.

In early December, the drill ship *JOIDES Resolution* will depart Colombo, Sri Lanka, and head for a spot in the southwestern Indian Ocean known as Atlantis Bank. There, it will lower a drill bit and try to screw it through 1.5 kilometres of rock, collecting a core sample as it goes. If all goes well, future expeditions — not yet scheduled or funded — will return and finalize the push into the mantle (see 'Deep understanding').

Normally, the crust–mantle boundary is

MORE ONLINE

PARIS CLIMATE TALKS

Daily news, commentary and video updates from the Paris climate talks go.nature.com/cslglg

MORE NEWS

- Improved molecular scissors for CRISPR go.nature.com/grgpai
- Genome editing: 7 facts about the latest technology go.nature.com/4rhrt6
- Billion-dollar boost for clean energy kicks off UN climate talks go.nature.com/fbhrzk

NATURE PODCAST

Solving the puzzle of fast radio bursts, finding missing baryons, and the future of the PhD nature.com/nature/podcast

NATURE VIDEO

J. BAYLOR ROBERTS/NATL GEOGRAPHIC/GETTY

thought to be marked by a feature known as the Mohorovičić discontinuity, or 'Moho', at which seismic waves change velocity. But at Atlantis Bank, the mantle is thought to bubble up as far as 2.5 kilometres above the Moho, making it easier to reach.

Reaching these deep-Earth frontiers "is one of the great scientific endeavours of the century", says Henry Dick, a geophysicist at the Woods Hole Oceanographic Institution in Massachusetts and co-leader of the expedition.

Beneath continents, the Moho lies 30–60 kilometres down. But beneath oceans it is close enough to be reached with ship-borne drilling equipment. In the drilling campaign — dubbed the Slow Spreading Ridge Moho, or 'SloMo' Project — Dick hopes to reach the crust–mantle transition at Atlantis Bank, then one day return with a state-of-the-art Japanese vessel to reach the Moho itself at a depth of 5 kilometres or more. Along the way, scientists aim to answer profound questions about the planet, such as how molten rock rises from the interior and cools to form fresh ocean crust, a surface that blankets three-fifths of Earth.

LONG-HELD DREAM

A hole that deep "would be the window into things we have never seen before", says Benoît Ildefonse, a geologist at the University of Montpellier in France.

Scientists first tried to reach the Moho in the middle of the twentieth century. In the 1960s, US scientists led 'Project Mohole', which drilled into the sea floor off Guadalupe Island, Mexico. The project reached a depth of just 183 metres before costs ballooned and Congress killed it. Still, Project Mohole gave birth to a series of scientific ocean-drilling programmes that have extracted cores from hundreds of locations around the world. These have revolutionized Earth science by retrieving sedimentary records that date back millions of years, offering clues to how continents pull apart and finding microbial life deep beneath the sea floor.

"We live on this Earth and we ought to know something about what happens beneath us," says Walter Munk, an oceanographer at the Scripps Institution of Oceanography in La Jolla, California, who conceived Project Mohole with colleagues over cocktails one evening in 1957. He is gratified by the success of scientific ocean drilling overall, but would still like to see the mantle breached.

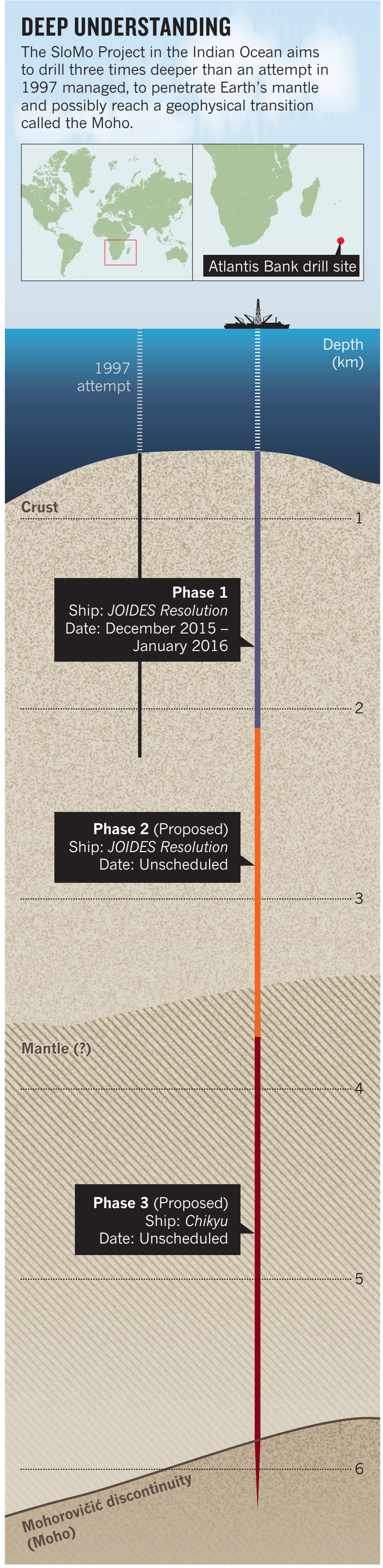
Expeditions have come close before. Between 2002 and 2011, four holes at a site in the eastern Pacific managed to reach fine-grained, brittle rock that geologists believe to be cooled magma sitting just above the Moho. But the drill could not punch through those tenacious layers. And in 2013, drillers at the nearby Hess Deep found themselves similarly limited by tough deep-crustal rocks (K. M. Gillis *et al. Nature* **505,** 204–207; 2014).

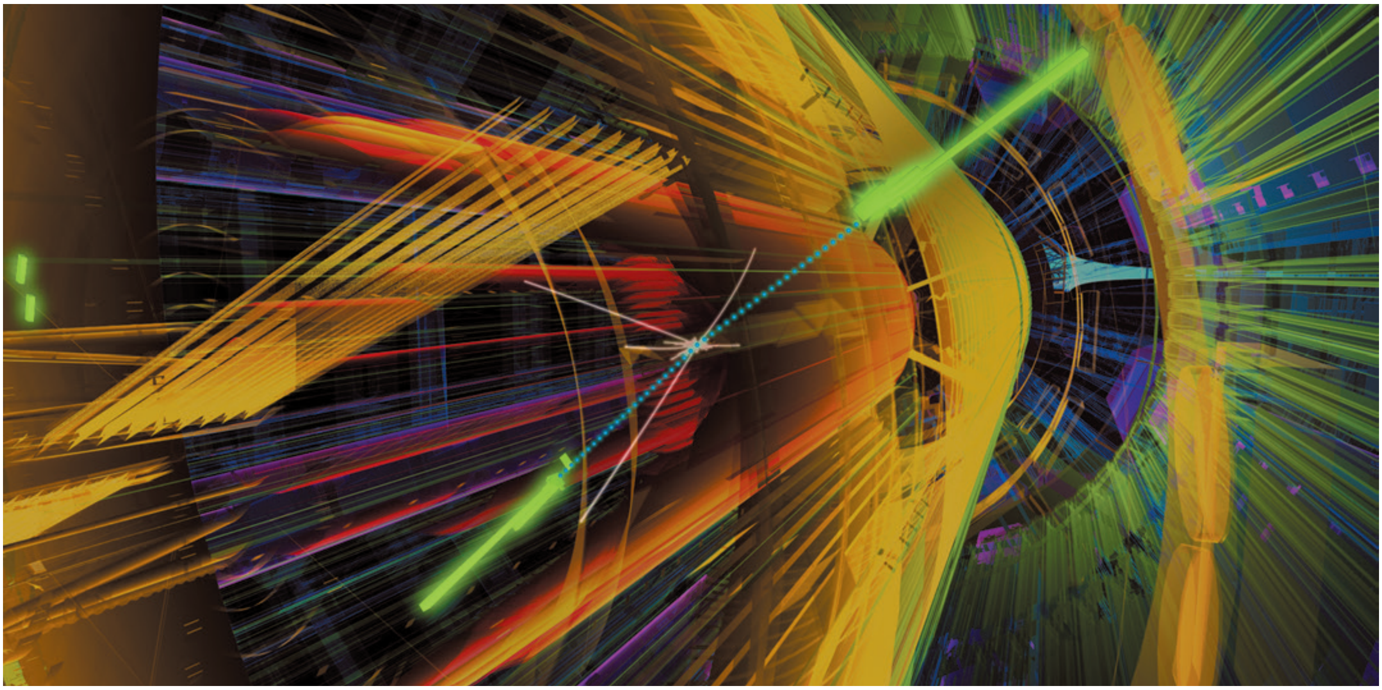
Dick and his colleagues are targeting the Indian Ocean ridge rather than the eastern Pacific because much smaller quantities of lava feed the sea floor there, so there is less hard rock to drill through. At Atlantis Bank, tectonic forces have lifted the sea floor to just 700 metres beneath the waves.

Dick knows that it is possible to reach his preliminary goal of 1.5 kilometres, because he has done it before. In 1997, he led an expedition to Atlantis Bank that got that deep before disaster struck: the pipe snapped off in high winds, corkscrewed down inside the hole and plugged it up. "We're going to make sure that doesn't happen this time," he says.

Along the way, researchers hope to explore not just geology, but biology, too. Geological mapping suggests that seawater may have percolated several kilometres deep at Atlantis Bank, triggering chemical reactions that turn the rock into a type known as serpentinite. These reactions generate methane, a gas that sub-sea-floor microbes often munch for energy. *JOIDES Resolution* scientists will be checking the rock cores for microorganisms, says Virginia Edgcomb, a microbiologist at Woods Hole who will be on the cruise.

SloMo's first phase runs until 30 January. If the drilling goes well, Dick hopes to return with the *JOIDES Resolution* to reach 3 kilometres. And after that, he and his colleagues hope to use the Japanese drill ship *Chikyu* in the project's third phase to drill all the way to the Moho. Launched a decade ago, *Chikyu* was meant to drill to the Moho in the western Pacific, but technical challenges and a lack of funding means that has not happened yet. With a capacity to drill as deep as 6 kilometres, *Chikyu* could finally allow geologists to realize their almost 60-year old dream. ■




CERN

Particle collisions at the Large Hadron Collider produce huge amounts of data, which algorithms are well placed to process.

PARTICLE PHYSICS

Artificial intelligence called in to tackle LHC data deluge

Algorithms could aid discovery at Large Hadron Collider, but raise transparency concerns.

BY DAVIDE CASTELVECCHI, GENEVA, SWITZERLAND

The next generation of particle-collider experiments will feature some of the world's most advanced thinking machines, if links now being forged between particle physicists and artificial intelligence (AI) researchers take off. Such machines could make discoveries with little human input — a prospect that makes some physicists queasy.

Driven by an eagerness to make discoveries and the knowledge that they will be hit with unmanageable volumes of data in ten years' time, physicists who work on the Large Hadron Collider (LHC), near Geneva, Switzerland, are enlisting the help of AI experts.

On 9–13 November, leading lights from both communities attended a workshop — the first of its kind — at which they discussed how advanced AI techniques could speed discoveries at the LHC. Particle physicists have "realized that they cannot do it alone", says Cécile Germain, a computer scientist at the University of Paris South in Orsay, who spoke at the workshop at CERN, the particle-physics lab that hosts the LHC.

Computer scientists are responding in droves. Last year, Germain helped to organize a competition to write programs that could 'discover' traces of the Higgs boson in a set of simulated data; it attracted submissions from more than 1,700 teams.

Particle physics is already no stranger to AI. In particular, when ATLAS and CMS, the LHC's two largest experiments, discovered the Higgs boson in 2012, they did so in part using machine learning — a form of AI that 'trains' algorithms to recognize patterns in data. The algorithms were primed using simulations of the debris from particle collisions, and learned to spot the patterns produced by the decay of rare Higgs particles among millions of more mundane events. They were then set to work on the real thing.

But in the near future, the experiments will need to get smarter at collecting their data, not just processing it. CMS and ATLAS each currently produces hundreds of millions of collisions per second, and uses quick and dirty criteria to ignore all but 1 in 1,000 events. Upgrades scheduled for 2025 mean that the number of collisions will grow 20-fold, and that the detectors will have to use more sophisticated methods to choose what they keep, says CMS physicist María Spiropulu of the California Institute of Technology in Pasadena, who helped to organize the CERN workshop. "We're going into the unknown," she says.

Inspiration could come from another LHC experiment, LHCb, which is dedicated to studying subtle asymmetries between particles and their antimatter counterparts. In preparation for the second, higher-energy run of the LHC, which began in April, the LHCb team programmed its detector to use machine learning to decide which data to keep.

LHCb is sensitive to tiny variations in temperature and pressure, so which data are interesting at any one time changes throughout the experiment — something that machine learning can adapt to in real time. "No one has done this before," says Vladimir Gligorov, an LHCb physicist at CERN who led the AI project.

Particle-physics experiments usually take months to recalibrate after an upgrade, says Gligorov. But within two weeks of the energy

upgrade, the detector had 'rediscovered' a particle called the J/Ψ meson — first found in 1974 by two separate US experiments, and later deemed worthy of a Nobel prize.

In the coming years, CMS and ATLAS are likely to follow in LHCb's footsteps, say Spiropulu and others, and will make the detector algorithms do more work in real time. "That will revolutionize how we do data analysis," says Spiropulu.

An increased reliance on AI decision-making will present new challenges. Unlike LHCb, which focuses mostly on finding known particles so they can be studied in detail, ATLAS and CMS are designed to discover new particles. The idea of throwing away data that could in principle contain huge discoveries, using criteria arrived at by algorithms in a non-transparent way, causes anxiety for many physicists, says Germain. Researchers will want to understand how the algorithms work and to ensure they are based on physics principles, she says. "It's a nightmare for them."

Proponents of the approach will also have to convince their colleagues to abandon tried-and-tested techniques, Gligorov says. "These are huge collaborations, so to get a new method approved, it takes the age of the Universe." LHCb has about 1,000 members; ATLAS and CMS have some 3,000 each.

Despite these challenges, the most hotly discussed issue at the workshop was whether and how particle physics should make use of even more sophisticated AI, in the form of a technique called deep learning. Basic machine-learning algorithms are trained with sample data such as images, and 'told' what each picture shows — a house versus a cat, say. But in deep learning, used by software such as Google Translate and Apple's voice-recognition system Siri, the computer typically receives no such supervision, and finds ways to categorize objects on its own.

Although they emphasized that they would not be comfortable handing over this level of control to an algorithm, several speakers at the CERN workshop discussed how deep learning could be applied to physics. Pierre Baldi, an AI researcher at the University of California, Irvine who has applied machine learning to various branches of science, described how he and his collaborators have done research suggesting that a deep-learning technique known as dark knowledge might aid — fittingly — in the search for dark matter.

Deep learning could even lead to the discovery of particles that no theorist has yet predicted, says CMS member Maurizio Pierini, a CERN staff physicist who co-hosted the workshop. "It could be an insurance policy, just in case the theorist who made the right prediction isn't born yet." ■

NEUROSCIENCE

Brain study seeks roots of suicide

A clinical trial will look at the neurological structure and function of people who have attempted suicide.

BY SARA REARDON

Suicide is a puzzle. Less than 10% of people with depression attempt suicide, and about 10% of those who kill themselves have never been diagnosed with any mental-health condition.

Now, a study is trying to determine what happens in the brain when a person attempts suicide, and what sets such people apart. The results could help researchers to understand whether suicide is driven by certain brain biologies — and is not just a symptom of a recognized mental disorder.

The project, which launched in November, will recruit 50 people who have attempted suicide in the 2 weeks before enrolling. Carlos Zarate, a psychiatrist at the US National Institute of Mental Health in Bethesda, Maryland, and his colleagues will compare these people's brain structure and function with those of 40 people who attempted suicide more than a year ago, 40 people with depression or anxiety who have never attempted suicide and a control group of 40 healthy people. In doing so, the researchers hope to elucidate the brain mechanisms associated with the impulse to kill oneself.

Zarate's team will also give ketamine, a psychoactive 'party drug', to the group that has recently attempted suicide. Ketamine, which is sometimes used to treat depression, can quickly arrest suicidal thoughts and behaviour — even in cases in which it does not affect other symptoms of depression[1]. The effect is known to last for about a week.

To some researchers, such findings suggest that ketamine affects brain circuits that are specific to suicidal thinking. But John Mann, a psychiatrist at Columbia University in New York City, says that abnormal brain chemistry and genetics could also predispose a person to attempt suicide in times of great stress, such as after a job loss. "They're part of the person, they're a trait," Mann says. "They just get more important when the person gets ill."

There is evidence that genetics influences a person's suicide risk. For instance, biological relatives of adopted children who kill themselves are several times more likely to take their own lives than the general population[2].

Fabrice Jollant, a psychiatrist at McGill University in Montreal, Canada, suggests that this genetic influence is related to impulsivity and flawed judgement, rather than to a specific mental illness. He has found that close relatives of people who killed themselves were more impulsive than a control group when playing a gambling game designed to test decision-making[3]. "It seems that this is something transmitted," Jollant says.

Other researchers are seeking biomarkers that would allow clinicians to spot the people most at risk of suicide. Alexander Niculescu, a psychiatrist at Indiana University in Indianapolis, and his colleagues have identified[4] a set of six genes whose expression is altered in the blood of people who have killed themselves. The team has found that combining these biomarkers with data from an app that tracks mood and risk factors can predict, with more than 90% accuracy, whether people with bipolar disorder or schizophrenia will eventually be hospitalized for a suicide attempt.

"In most clinical trials, people at high risk of suicide are excluded, so we don't know how to treat them."

Researchers hope that a better understanding of the biology that underlies suicide will lead to more effective treatments for suicidal impulses. But studies such as Zarate's present difficult logistical and ethical challenges. Researchers must consider whether a person who has just attempted suicide can make informed decisions about whether to participate in research.

Those who study suicidal people say that they treat them with special care — and that the overall benefits of such studies outweigh any risks. "In most clinical trials, people at high risk of suicide are excluded, so we don't know how to treat them," Jollant says. "We need to assess this population, not just say 'exclude them from trials.'" ■

1. Ballard, E. D. *et al. J. Psychiatr. Res.* **58,** 161–166 (2014).
2. Brent, D. A. & Mann, J. J. *Am. J. Med. Genet. C Semin. Med. Genet.* **133C,** 13–24 (2005).
3. Hoehne, A. *et al. J. Psychiatr. Res.* **68,** 192–197 (2015).
4. Niculescu, A. B. *et al. Mol. Psychiatr.* **20,** 1266–1285 (2015).

POLICY

UK science budget goes up

Celebrations as spending set to rise with inflation.

BY ELIZABETH GIBNEY

UK scientists' worst funding fears have not come to pass. The country's science budget will rise slightly in the coming years, Chancellor of the Exchequer George Osborne said in a much-anticipated government spending review.

Ahead of the review, scientists had braced for the possibility that spending would remain flat — as it has for the past five years — and continue to be whittled away by inflation, or even be cut. But speaking in the House of Commons on 25 November, Osborne announced that the £4.7-billion (US$7.1-billion) science budget will now rise with inflation. This would amount to an extra £500 million for science annually by the end of the decade, according to the Treasury. Osborne also committed to increasing the £1.1-billion annual budget for science infrastructure to £1.2 billion a year by 2020–21.

Scientists' initial reaction was relief. "If the science budget is really protected in real terms, then that is good news," says Lee Cronin, a chemist at the University of Glasgow. Naomi Weir, acting director of the Campaign for Science and Engineering in London, said in a statement: "This announcement is great news for the UK."

However, Cronin and others noted that there is work to be done to reverse the damage caused by the flat budget. Although the increase in infrastructure spending will be helpful, Cronin adds that it needs to be "used to help replace essential equipment and provide the upgrades needed urgently, rather than just fund shiny new projects".

While acknowledging that the outcome could have been much worse, Jenny Rohn, who chairs the UK lobby group Science is Vital, highlighted that the science budget is smaller in real terms in 2015 than it was in 2010, owing to erosion by inflation.

The science budget will also have to cover a new Global Challenges research fund, aimed at addressing the problems faced by developing countries.

Osborne announced that the government would implement the recommendations of a review by geneticist Paul Nurse to create Research UK, a new umbrella body to oversee the seven research councils that distribute most of the science budget. ■

France's President François Hollande attends a national tribute to the victims of the Paris terrorist attacks.

PASCAL LE SEGRETAIN/GETTY

TERRORISM

Why Europeans turn to jihad

Terrorism is tough to study, but researchers have gleaned insights from the current generation of Islamist extremists.

BY DECLAN BUTLER

In the wake of the terrorist attacks in Paris on 13 November that left 130 dead and more than 350 wounded, Alain Fuchs, president of the French National Centre for Scientific Research (CNRS), announced a fresh call for proposals for research on terrorism. Acknowledging that any effort with no immediate effect may seem "derisory", Fuchs said that science can help to open up avenues of analysis.

The Islamist terror group ISIS also carried out deadly attacks this year in Tunisia, Lebanon, Bangladesh and other countries, and downed a Russian airliner in the Sinai Peninsula. But as thousands of Europeans have left to join Islamist groups in conflict zones, and are at risk of returning home trained to carry out further attacks, the continent is on edge.

Terrorism researchers are trying to understand how young people in Europe become radicalized, by looking for clues in the life histories of those who have committed or planned terrorist acts in recent years, left the continent to join ISIS, or are suspected of wanting to become jihadists. A mixture of sociologists, political scientists, anthropologists and psychologists, such researchers are drawing on information generated by police, judicial inquiries and the media, and, in some cases, on interviews. They also study factors at play in prisons and socially-deprived areas. Some of their insights are summarized here.

Religion is not the trigger. The rise of jihad in Europe has led to an assumption that there is a radicalization of Muslims more generally across the continent. Yet research suggests that most extremists are either people who returned suddenly to Islam or converts with no Islamic background, says Olivier Roy, who specializes in political Islam and the Middle East at Italy's European University Institute near Florence — and as many as one in four French jihadists is a convert. Roy summarized the latest research at a conference organized in Mainz on 18–19 November by the German Federal Criminal Police Office.

Violent extremism emerges first, with a religious justification tagged on after, adds Rik Coolsaet, head of political science at Ghent University in Belgium, who studies jihadis and foreign policy. He notes that two young British men who were jailed last year on terrorism offences after fighting in Syria had

earlier ordered online the books *Islam for Dummies* and *The Koran for Dummies*.

Resentment is the common ground. It is difficult to make generalizations about how people become radicalized in Europe. At the Mainz conference, Roy said that many extremists come from broken families or deprived areas, lack education and are unemployed. A smaller number are well educated, have held jobs and have middle-class lifestyles. Some are in stable relationships and have young children. The characteristics that extremists seem to share are resentment directed at society and a narcissistic need for recognition that leaves them open to a narrative of violent glory, said Roy.

Social factors can contribute to such frustrations, according to Farhad Khosrokhavar, a CNRS researcher who works at the School for Advanced Studies in Social Sciences in Paris. Almost all European extremists and terrorists are second- and third-generation immigrants, whom Khosrokhavar says are often "stigmatized, rejected and treated as second-class citizens". However, since about 2013, the profile of those leaving to fight in Syria has included a much larger proportion of middle-class youth than in previous generations, he says.

Terrorism breeds in prisons. The link between terrorism and prison was highlighted this year. The three terrorists involved in the January attack in Paris on the satirical publication *Charlie Hebdo* and a kosher supermarket, as well as some of the 13 November attackers, had all done time.

Many French terrorists have a history of petty crime that landed them in prison. Stays there often proved seminal experiences on their path to radicalization, says Khosrokhavar, who spent several years interviewing some 160 staff and inmates at 4 large French prisons, including 15 inmates sentenced for terrorism offences. He says that prisoners often come under the influence of — and form lasting bonds with — radical Islamists and terrorist networks.

Since about 2013, the profile of those leaving to fight in Syria has included a much larger proportion of middle-class youth.

'Entrepreneurs' drive terrorism. Most of those who get involved in jihadi terrorism in Europe are "misfits and drifters" — people who joined militant networks during life crises or through friends and relatives on the inside, says Petter Nesser, a terrorism researcher at the Norwegian Defence Research Establishment in Kjeller.

But he says that the key actors in terrorist activity are a much smaller number of "entrepreneurs". These seasoned, ideologically driven activists are part of transnational terrorist webs linked both to extremist groups throughout Europe and to armed groups in conflict zones. They are the ones who bring structure and organization to the disaffected majority, through recruitment and indoctrination.

Molenbeek isn't the terrorist capital of Europe. Several of the terrorists involved in the latest Paris attacks, and the perpetrators of previous attacks in Europe, had lived in the Molenbeek district of Brussels, which has a large Muslim community, mostly of Moroccan descent. This has led some politicians and media outlets to label it Europe's terrorism capital — and to blame factors such as social deprivation or an apparent lack of integration of Muslims.

"This is misleading," says Nesser. Jihadi hot spots have emerged across Europe in environments ranging from poor suburbs, to universities and schools, to prisons. The key ingredient in the spread of jihadism in any location is a critical mass of jihadist entrepreneurs, he says.

A focus on Molenbeek obscures the fact that European jihadism is transnational, Nesser says, and that its main drivers are armed conflicts and militant groups involved in those conflicts. He adds: "It is also unfair and stigmatizing towards the inhabitants of this Belgian suburb." ■ SEE EDITORIAL P.7

How to build a better PhD

There are too many PhD students for too few academic jobs — but with imagination, the problem could be solved.

BY JULIE GOULD

"Since 1977, we've been recommending that graduate departments partake in birth control, but no one has been listening," said Paula Stephan to more than 200 postdocs and PhD students at a symposium in Boston, Massachusetts, in October this year.

Stephan is a renowned labour economist at Georgia State University in Atlanta who has spent much of her career trying to understand the relationships between economics and science, particularly biomedical science. And the symposium, 'Future of Research', discussed the issue to which Stephan finds so many people deaf: the academic research system is generating progeny at a startling rate. In biomedicine, said Stephan. "We are definitely producing many more PhDs than there is demand for them in research positions."

The numbers show newly minted PhD students flooding out of the academic pipeline. In 2003, 21,343 science graduate students in

ILLUSTRATION BY OLIVER MUNDAY

the United States received a doctorate. By 2013, this had increased by almost 41% — and the life sciences showed the greatest growth. That trend is mirrored elsewhere. According to a 2014 report looking at the 34 countries that make up the Organisation for Economic Co-operation and Development, the proportion of people who leave tertiary education with a doctorate has doubled from 0.8% to 1.6% over the past 17 years.

Not all of these students want to pursue academic careers — but many do, and they find it tough because there has been no equivalent growth in secure academic positions. The growing gap between the numbers of PhD graduates and available jobs has attracted particular attention in the United States, where students increasingly end up stuck in lengthy, insecure postdoctoral research positions (see *Nature* **520,** 144–147; 2015). Although the unemployment rate for people with science doctorates is relatively low, in 2013 some 42% of US life-sciences PhD students graduated without a job commitment of any kind, up from 28% a decade earlier. "But still students continue to enrol in PhD programmes," Stephan wrote in her 2012 book *How Economics Shapes Science*. "Why? Why, given such bleak job prospects, do people continue to come to graduate school?"

One reason is that there is little institutional incentive to turn them away. Faculty members rely on cheap PhD students and postdocs because they are trying to get the most science out of stretched grants. Universities, in turn, know that PhD students help faculty members to produce the world-class research on which their reputations rest. "The biomedical research system is structured around a large workforce of graduate students and postdocs," says Michael Teitelbaum, a labour economist at Harvard Law School in Cambridge, Massachusetts. "Many find it awkward to talk about change."

"We need to transform graduate education within five years. It's imperative."

But there are signs that the issue is becoming less taboo. In September, a group of high-profile US scientists (Harold Varmus, Marc Kirschner, Shirley Tilghman and Bruce Alberts, colloquially known as 'the Quartet') launched Rescuing Biomedical Research, a website where scientists can make recommendations on how to 'fix' different aspects of the broken biomedical research system in the United States — the PhD among them. "How can we improve graduate education so as to produce a more effective scientific workforce, while also reducing the ever-expanding PhD workforce in search of biomedical research careers?" the site asks.

Nature put a similar question to 33 PhD students, scientists, postdocs and labour economists and uncovered a range of opinions on how to build a better PhD system, from small adjustments to major overhauls. All agreed on one thing: change is urgent. "Academia really is going to have to be dragged kicking and screaming into the twenty-first century," says Gary McDowell, a postdoctoral fellow at Tufts University in Medford, Massachusetts, and a leader of the group behind the Future of Research symposium. The renovation needs to happen now, says Jon Lorsch, director of the US National Institute of General Medical Sciences in Bethesda, Maryland. "We need to transform graduate education within five years. It's imperative. There's a lot at stake for scientists, and hence for science."

TRACK THE PHD

One place to begin is with hard facts: show prospective students and supervisors data on trainees' chances of moving into academic research or other careers. Prospective students "aren't thinking strategically about what they really want to do or what they're best suited for", says Patricia Labosky, a programme director for scientific training at the US National Institutes of Health (NIH) in Bethesda, Maryland.

A 2015 *Nature* survey of more than 3,400 science graduate students around the world suggested that many were overly optimistic about their chances in academia. About 78% of respondents said that they were "likely" or "very likely" to follow an academic career, and 51% thought that they would land some type of permanent job in one to three years. In reality, only about 26% of PhD students in the United States move into tenured or tenure-track positions, and getting there can take much longer than this (see 'Ups and downs of PhDs').

But although some data exist about career paths, there are key gaps relating to the range of job opportunities, earnings, time spent as a postdoc and long-term career trajectories, says Julia Lane, an economist at New York University. A January report on post-PhD careers by the US Council of Graduate Schools in Washington DC found that there are no standardized ways to collect information on graduates after they have left their educational institution; only around one-third of universities in the United States and Canada formally compile such data.

In October, Stanford University in California published the results of a major effort to track graduates either 5 or 10 years after their PhD. It showed that the number of bioscience PhD students progressing to postdoctoral positions had dropped from 41% to 31% in the more recent graduate group, and that many were moving into business, government or non-profit positions. This probably reflects the growing bottleneck in academic jobs and booming opportunities in business.

Lane is leading a more comprehensive effort to track career outcomes in research called UMETRICS, which is based at the University of Michigan in Ann Arbor. By combining anonymized human-resource and administrative data from universities with US Census Bureau data on earnings, places of work and job titles, UMETRICS will be able to produce campus-level reports on the career outcomes of graduate students. A student interested in a chemistry PhD, for example, could scan a campus report and see what previous graduates went on to do, where they went and how much they earn. It will take several years before the first data sets are released, Lane says — but when they are, "the students opting in to graduate schools will go in with eyes wide open".

REVAMP THE PHD

Many PhD students enjoy the intellectual freedom of a PhD for a few years and then successfully move on to other things. But a lot of students want more preparation and training for that step — such as building skills in management, budgeting or negotiation. "Apparently, you have to learn these things somewhere on the side, since you are supposed to spend all your time as a PhD and postdoc doing research," says Joanna Klementowicz, a postdoc at the University of California, San Francisco (UCSF).

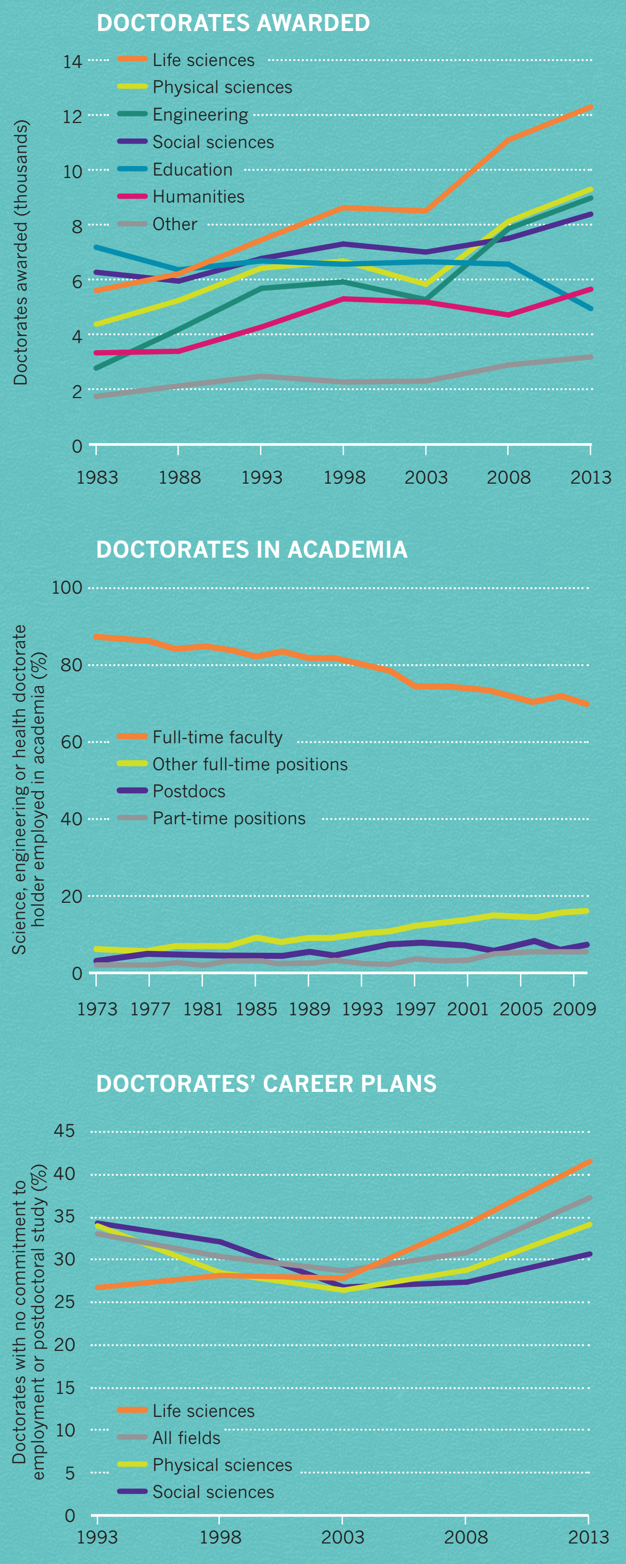
The current graduate education system in many countries is based on an apprenticeship model, wherein lab heads train younger researchers in the craft of research. This system has been prominent since the 1800s, when the first 'modern' PhD was awarded by the University of Berlin. Although the scientific enterprise has changed dramatically since then, the PhD system has not.

Modernizing the PhD could improve training in areas of research ranging from reproducibility to experimental design and entrepreneurship. It could also help to solve the bottleneck problem by equipping doctorate holders with soft skills that make them more employable wherever they go. "We need to tailor graduate education to meet the needs of students without violating what it means to be a scientist," says Alan Leshner, chief executive emeritus of the American Association for the Advancement of Science in Washington DC.

Some funding bodies and research institutions have already taken this on board. In 2013, the NIH started the Broadening Experiences

UPS AND DOWNS OF PHDS

The number of students in the United States who graduate with a doctorate has increased, with the most rapid rise in life-sciences degrees. The proportion of PhDs in permanent academic positions is falling, and the number graduating with no job or postdoc lined up is on the rise.



SOURCE: NATIONAL CENTER FOR SCIENCE AND ENGINEERING STATISTICS

in Scientific Training (BEST) initiative — a US$3.7-million programme that is designed to improve training for biomedical PhDs and postdocs. "We got a lot of feedback from [employers] that the graduates weren't ready for careers outside of academia," says Labosky, who heads the programme.

At UCSF, PhD students on the BEST programme spend nine months training in areas such as management, interviewing and networking, and are put into groups that work together to explore career objectives. "The programme made me practical: I learned to look out for what I can apply for, what my skills were matched to and what people with a PhD like mine go on to do," says Klementowicz, who took the programme as a postdoc.

Some scientists would like to see particular emphasis put on teamwork to reflect the increasingly collaborative nature of research. David Golan, dean of graduate education at Harvard Medical School in Boston, Massachusetts, is considering how to ingrain teamwork more deeply into the graduate-school experience. "We have toyed with the idea of having students form a team before they apply to grad school," he says. They might then be given a project to work on together throughout their training — and perhaps even be examined together.

SPLIT THE PHD

There may be too many PhD graduates for academia, but there is plenty of demand for highly educated, scientifically minded workers elsewhere. So some scientists propose that the PhD should be split into two: one for future academics and a second to train those who would like in-depth science education for use in other careers.

Biologist Anthony Hyman, director of the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany, is one of those who thinks that a split PhD might work. Students in the academic-track PhD would focus on blue-skies research and discovery, he says. A vocational PhD would be more structured and directed towards specific careers in areas such as radiography, machine learning or mouse-model development.

A similar concept already exists in engineering: students in the United Kingdom, the United States, France and Germany can choose to study for either an academic-style PhD in engineering or a doctorate in engineering (EngD), which is designed with industrial careers in mind and often involves a supervisor in industry alongside one in academia. David Stanley, who manages an EngD programme that focuses on nuclear engineering at the University of Manchester, UK, says that the programme is aimed at supplying industry with employees. "Graduates with an EngD are highly valued in industry, more than those with PhDs, because of their extended training," he says.

Elsewhere, industrial PhDs are taking shape in the biomedical sciences. One of the oldest government-organized industrial PhD schemes is run by Innovation Fund Denmark, which supports students who are simultaneously enrolled at a Danish university and employed (and paid) by a private-sector company. Melanie Sinche, director of education at the Jackson Laboratory for Genomic Medicine in Farmington, Connecticut, is enthusiastic about the idea of a vocational PhD at her institute, where it might fulfil a need for more expert computational biologists. "The number of people qualified to do this is small, and there are lots of employers competing for this small pool of candidates," she says.

But the split PhD could face challenges if the two tracks are valued in different ways: academics could view a vocational PhD as second-class, whereas tech companies could view an academic PhD as too abstruse for the real world. That could end up limiting the career options of doctorates rather than broadening them, says Hyman. Stanley counters that EngD students do not have that problem. "A couple of students a year find their way back into academia to conduct research," he says.

➲ NATURE.COM
Listen to a podcast on the future of the PhD at:
go.nature.com/i8yh8f

SKIP THE PHD

Some scientists call for more drastic measures — cutting down the number of people who pursue a PhD.

Siphoning off more students into master's programmes is one way to reduce PhD numbers, says Bruce Alberts, professor of biochemistry and biophysics in the department of medicine at UCSF. A master's can offer advanced scientific training that is sufficient for many careers, as well as a taste of research, in one or two years rather than the four or five eaten up by a typical PhD. "In an ideal world, everyone would go in for a master's," Alberts says.

Master's degrees are already common across Europe. In the Netherlands, students are required to complete a master's before embarking on a PhD. "There are many who don't want to be in academia who leave with a master's to work in government institutions, companies, in publishing," says Frank Miedema, professor and head of immunology at the University Medical Center Utrecht in the Netherlands. "And a master's is not considered a failure for those who can't make it to a PhD."

Victoria Evans graduated with a master's degree in astrophysics from Cardiff University, UK, in 2012. "The research project in the master's gave me an insight into what a PhD project would be like," she says, "and I came to the conclusion that it wasn't what I wanted to do." She now works as a nuclear-safety engineer for EDF Energy on the west coast of Scotland. "The problem-solving and analytical skills that I learned during my master's were more than sufficient for me to work in this field."

"In an ideal world, everyone would go in for a master's."

In the United States, the science master's has often had a lower status than the PhD — but universities are now launching more of them. Between 2000 and 2011, the number of science and engineering master's degrees available increased by 57%, compared with a 38% increase in doctoral degrees, according to the US National Science Foundation. Part of that growth has been in the professional science master's degree, a programme developed in the late 1990s as a graduate degree that would simultaneously develop scientific and workplace skills. Last year, Harvard Medical School introduced a two-year master's in immunology aimed at students who want additional classroom and research experience to help them decide whether to continue on to a PhD or MD, or to transition to industry.

But master's programmes are no panacea. Unlike most doctoral students, master's students in the United States and Europe are often required to pay for their tuition, and that could dissuade many from signing up. "This does create a social access problem," said neuroscientist Eve Marder of Brandeis University in Waltham, Massachusetts, at last month's Future of Research meeting.

CUT THE PHDS

Labour economists have been advocating for a reduction in the number of graduate students who enter biomedical sciences for several decades. Yet there is enormous resistance to change. That's what the Quartet found, when it proposed gradually reducing the numbers of PhD students as part of its efforts to rescue biomedical research. "This idea has had the most opposition from our colleagues," says Alberts. Faculty members and research institutions may be especially reluctant to give up the cheap workers who power their research when government funding for biomedicine has fallen, as it has in the United States for the past decade or so. And some scientists argue that fewer PhD graduates would be a loss to science and society as a whole. "The draconian measures of restricting access to graduate school is detrimental to science," said Marder at the Future of Research meeting. "It means we would restrict the imagination in our workforce."

Cuts to PhD programmes haven't gone down well. When the Canadian Institutes of Health Research cancelled its 30-year-old MD/PhD programme earlier this year owing to budget tightening, academics and students reacted with horror. But other fields regulate the flow of students into courses to match supply to demand. The American Bar Association, which oversees the legal system in the United States, attempts to regulate the number of qualified lawyers by exerting strict control over the number of law schools. And bar associations set fiendishly difficult examinations for would-be lawyers to get into law school in the first place.

Stiffer entrance assessments for those who want to pursue a PhD could cut down entrant numbers — if the right criteria can be found. In the United States, Graduate Records Examinations (GREs) are used as a way of selecting entrants for graduate school, but the system is hardly perfect: one survey showed that 37% of US biology PhD students drop out before completing their degree. When Orion Weiner, a molecular biologist at UCSF, did a small, retrospective study of graduate students admitted onto one of his university's biology PhD programmes, he found that previous experience in research and the subject-specific GRE results (but not the analytical, verbal or quantitative elements) were good indicators of future success in graduate school.

A broader entrance assessment could look at students' experience in communication, management, teamwork and career goals. That could be used to filter students with a passion for academic or industrial research towards PhD programmes and send others into a master's or other types of training, says Bill Lindstaedt, executive director for career advancement at UCSF.

Stephan believes that funding bodies should have a major role in limiting the number of biomedical PhD places to better match supply and demand, and she also proposes that students should contribute to their training costs. "When we have to pay something out of pocket, we think a little more clearly about whether that is a good fit for us," she says. Such ideas may be controversial — but many people say that they have to be considered.

At the heart of the problem, say scientists, is that the community is not discussing the PhD problem enough. "There is a reluctance from supervisors to tell undergrads and grad students the reality of the system," says postdoc McDowell. "The misinformation exists because the system is worried about deflecting smart people from entering." Although principal investigators acknowledge the difficulty of securing an academic position, the system worked for them and so it is tempting to tell students that they can do it too — just another experiment, another publication or another year, and you'll get there.

Grass-roots groups such as Future of Research are calling attention to the issue, as are efforts such as Rescuing Biomedical Research. Meanwhile, some experts say that the onus falls partly on prospective and current PhD students to make sure their eyes are open. They should arm themselves with as much information as possible, says Labosky, so that "they are aware of their alternative options and can make plans".

Stephan does see some prospect that her call for PhD birth control will be heard. She says that change might happen naturally, as more information becomes available on career outcomes, and that flat funding streams could prevent further growth in biomedical PhDs. "Individuals might become less focused on PhD production, and universities and faculty are more likely to pay attention to these recommendations."

Teitelbaum, for his part, does not favour a large cut in biomedical PhDs, and instead prefers a more considered approach. "Find out why people start PhDs and what they think their career prospects are from the very beginning," he says. "Like ballet dancers or actors, if they chose to take it on knowing their chances of becoming a successful professor, then let them carry on." ■ SEE EDITORIAL P.7 AND CAREERS P.155

Julie Gould *is an editor for* Naturejobs.

A. CHEZIÈRE/BEL/EMSE

THE BODY ELECTRIC

RESEARCHERS WANT TO WIRE THE HUMAN BODY WITH SENSORS THAT COULD HARVEST REAMS OF DATA — AND TRANSFORM HEALTH CARE.

BY ELIZABETH GIBNEY

Göran Gustafsson looks at people and thinks of cars — the ageing models that rolled off assembly lines a few decades ago. Today, says Gustafsson, cars are packed with cutting-edge sensors, computers and sophisticated communications systems that warn of problems when they are still easy to fix, which is why modern vehicles rarely surprise their drivers with catastrophic breakdowns.

"Why don't we have a similar vision for our bodies?" wonders Gustafsson, an engineer whose team at the Swedish electronics company Acreo, based in Kista, is one of many around the world trying to make such a vision possible. Instead of letting health problems go undetected until a person ends up in hospital — the medical equivalent of a roadside breakdown — these teams foresee a future in which humans are wired up like cars, with sensors that form a similar early-warning system.

Working with researchers at Linköping University in Sweden, Gustafsson's team has developed skin-surface and implanted sensors, as well as an in-body intranet that can link devices while keeping them private. Other groups are developing technologies ranging from skin patches that sense arterial stiffening — a signal of a looming heart attack — to devices that detect epileptic fits and automatically deliver drugs directly to affected areas of the brain.

These next-generation devices are designed to function alongside tissue, rather than be isolated from it like most pacemakers and other electronic devices already used in the body. But making this integration work is no easy feat, especially for materials scientists, who must shrink circuits radically, make flexible and stretchable electronics that are imperceptible to tissue, and find innovative ways to create interfaces with the body. Achieving Gustafsson's vision — in which devices monitor and treat the body day in, day out — will also require both new power sources and new ways of transmitting information.

Still, the potential to improve health care substantially while reducing its costs has drawn both researchers and physicians to the challenge,

Surface sensors need to be as flexible and stretchy as the skin they are mounted on.

says John Rogers, a materials scientist at the University of Illinois at Urbana–Champaign. "I haven't found any clinical folks who say 'That's pie in the sky, come back to me in 20 years,'" he says. "They say, 'Wow, that's cool. Here are three ways we can use it today, and how do we get started on a collaboration?'."

Sensors woven into the body are a natural extension of handheld smartphones and wearable devices, says Rogers (see *Nature* **525,** 22–24; 2015). "I think electronics is coming at you," he says. "It's migrating closer and closer and I think it's a very natural thing to imagine that they will eventually become intimately integrated with the body."

SKIN DEEP

The first step beyond wearables will be wireless sensors mounted directly on the skin, where they can pick up a host of vital signs, including temperature, pulse and breathing rate. Unfortunately, says Rogers, "biology involves bending, stretching and swelling", which makes conventional electronics built from stiff silicon wafers a very poor choice for such sensors.

His team has developed 'epidermal electronics': flexible, biodegradable stick-on patches that are crammed with sensors but almost imperceptible to the user. Attached like temporary tattoos, the patches use normal silicon electronics, but thinned down and transferred to a flexible backing using a rubber stamp[1]. The patches draw power either from nearby magnetic fields or by harvesting radio waves, using S-shaped wires and antennas designed to stretch, twist and bend. "They adopt a wavy kind of geometry, so when you stretch, the wave shapes can change, like accordion bellows," says Rogers.

Rogers has co-founded a spin-off company — MC10, based in Lexington, Massachusetts — that next year will start marketing versions of the device as BioStamps: temporary patches that measure heart electrical activity, hydration, body temperature and exposure to ultraviolet light. The patches will be available to consumers first, says Rogers, but his real target is medicine. Results are expected soon from a trial at the neonatal intensive-care unit at Carle Foundation Hospital in Urbana, where doctors are using the patches to monitor the vital signs of newborn babies without the need for intrusive cables and scanners. MC10 is also collaborating with Brussels-based pharmaceutical company UCB on tests of a patch that monitors tremors in people with Parkinson's disease, to track their illness and whether they are taking their medication.

Rogers' patches are relatively small, but at the University of Tokyo, engineer Takao Someya has created a sensor-laden electronic skin that can be made in much larger pieces[2]. His latest film is just 1 micrometre thick, and so light that it floats like a feather, yet it is robust enough to cope with the stretching and crumpling needed to flex with an elbow or knee. It can provide readouts on temperature — heat in a wound can signal infection — moisture, pulse and oxygen concentration in the blood. Someya achieves this by ditching silicon altogether, and instead using inherently soft organic components made of carbon-based polymers and other materials. Organic circuits can be printed onto a plastic film, making them cheap and easy to produce in large quantities. And they are versatile: they work in both high-temperature and water-based environments.

Skin also inspires Zhenan Bao, an engineer at Stanford University in California. Her team creates thin pressure sensors by sandwiching micrometre-scale rubber pyramids between films[3]. Even a slight touch will compress the pyramids' tips, changing how electric current flows between the films. The sensors can be used in heart monitors that track how fast pressure waves pass through arteries. This can reveal increased stiffness in the vessels — a predictor of heart attacks. Last year, the US Food and Drug Administration approved a wireless pressure sensor that can be implanted inside the hearts of people with advanced heart disease; Bao's device could do a similar job from the surface of the skin.

As useful as skin-mounted patches might be, much more information is available deeper in the body. "There's a reason why at the hospital, they draw your blood," says Michael Strano, a chemical engineer at the Massachusetts Institute of Technology (MIT) in Cambridge. "There are markers in blood that are exquisitely good at predicting disease."

But delving deeper brings fresh challenges. Ideally, says Strano, sensors under the skin should be not only non-toxic, but also stable enough to function inside the body for years at a time if need be, and biocompatible — meaning that they don't trigger the body's immune response. Yet most current devices fall short on one score or another. For example, sensors that detect chemical signals in the blood called biomarkers often use biological materials that degrade very quickly. This is a severe limitation for the advanced, real-time sensors that are currently used to monitor glucose in people with diabetes, says Strano: the devices detect glucose with an enzyme reaction that produces hydrogen peroxide. This degrades the sensors so quickly that they must be replaced within weeks.

To get around that, Strano's lab has developed synthetic, long-lived detector materials that can be mixed with a water-based gel and injected under the skin like a tattoo. The 'ink' for this tattoo consists of carbon nanotubes coated with dangling polymer strands, which have a lock-and-key chemical structure that recognizes biomarkers by dictating which molecules can dock with them[4]. When biomarkers bind to the polymer, they subtly change the optical properties of the nanotube: shine a light on the tattoo, and a glow reveals the presence of the biomarker.

Strano and his team have developed carbon-nanotube sensors to monitor nitric oxide in blood[5] — an inflammatory marker that can indicate infection or even cancer — and are working on glucose and cortisol, a stress biomarker that may prove useful for monitoring post-traumatic stress disorder and anxiety disorders. The nitric oxide sensor worked for 400 days in mice, which to Strano's knowledge is the longest any implanted chemical sensor has been in place, and did so without provoking any immune response. For many other kinds of device, the jury is still out. "For electronic materials, especially plastic-based and organics, it's still unknown what their long-term effects are," says Bao.

> "I HAVEN'T FOUND ANY CLINICAL FOLKS WHO SAY, **'THAT'S PIE IN THE SKY.' THEY SAY, 'WOW, THAT'S COOL.'"**

Now Strano is starting work with MIT engineer Daniel Anderson on devices that could combine sensors with drug-delivery systems. They hope to adapt microchips pioneered by fellow MIT engineer Robert Langer to respond to a range of triggers by releasing the appropriate drugs, encased in polymer capsules. The first human trial of a drug-delivering 'pharmacy on a chip' — without the sensors — was in 2012, in eight women with osteoporosis[6].

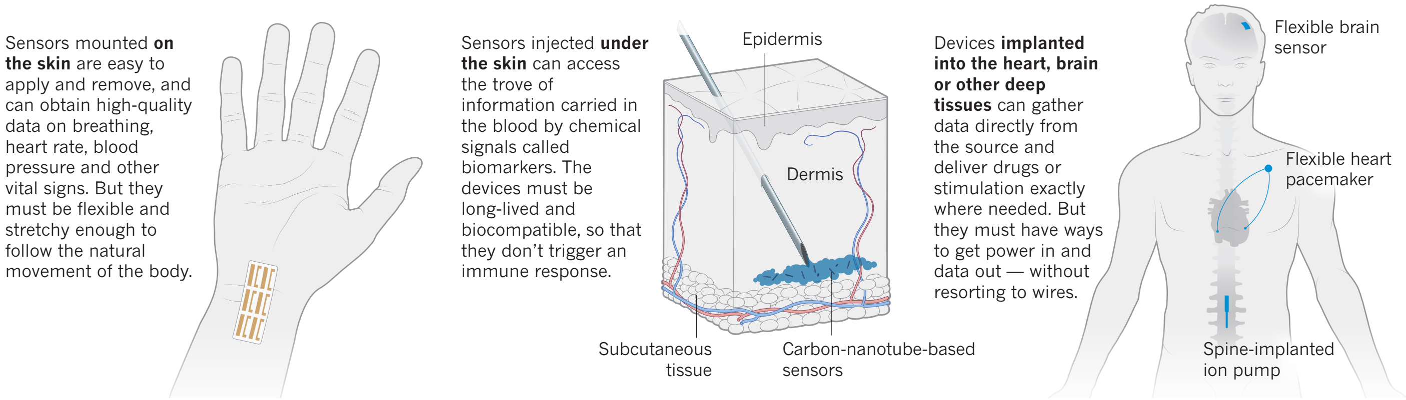
It will be a long time before such devices can be used to detect diseases reliably and treat them automatically, except perhaps for diabetes, which has been extensively studied. Strano's devices are good at binding only with their target molecules, but big questions remain about what fluctuations in biomarker signals actually mean in terms of health, he says. His team is modelling biomarkers in the body, to help to decide where the sensor needs to be and how quickly it should to react to give useful information. "Often you need to rely on many different sensory parameters to make a decision. It's not enough that one chemical is over-expressed," says Magnus Berggren, an electronic engineer at Linköping University who is collaborating with Gustafsson.

MOVING TARGET

Some researchers' targets lie still deeper in the body, and for them, flexibility and biocompatibility are even more important. If a rigid sensor rubs against a moving organ such as the heart or the brain, in which the cells shift slight as the animal breathes, the body will quickly surround it with a wall of scar tissue. And if sensors move relative to the organ, the results will be unreliable in any case.

WIRED FOR LIFE

Sensors woven into the body could alert people to medical problems before they become seriously ill — if the devices can overcome some daunting challenges.



Bioelectronics engineer George Malliaras at the École Nationale Supérieure des Mines de Saint-Étienne in Gardanne, France, and his colleagues are among those developing flexible replacements for the relatively rigid sensors currently used to track distinctive electrical patterns in the brains of people with epilepsy or Parkinson's disease. Made of organic, conducting polymers, these flexible electronics respond to chemical signals — the flow of ions that generates the electrical patterns. This not only increases sensitivity, but also lets researchers "interface with biology in a wholly different fashion", he says.

The team's latest device, tested in rats as well as in two humans undergoing surgery for epilepsy[7], has detected the firing of individual neurons, says Malliaras. And if the process is reversed, he adds, the sensor can be used to deliver drugs. Devices known as organic electronic ion pumps respond to an applied voltage by forcing drugs — small charged particles — out of a reservoir. Working with the group at Linköping University and the French National Institute of Health and Medical Research in Marseilles, Malliaras's team is coupling his epilepsy sensor to an ion pump that responds to seizures by releasing epilepsy drugs into the correct part of the brain[8]. Berggren and the Linköping team have used a similar technique to develop a 'pacemaker for pain' that delivers analgesics directly to the spinal cord[9].

KEEP IT GOING

Any electrical device is limited by its need for power. Devices that sit on or near the skin can incorporate antennas that harvest power wirelessly — as long as an external source is nearby. But sensors deeper in the body often have to rely on batteries, which are bulky and need replacing. And some, such as Berggren's pain-relief pump, need to have wires threaded through the overlying tissues — an arrangement that is both cumbersome and a potential route for infection (see 'Wired for life').

To get around such problems, Zhong Lin Wang, a nanoscientist at the Georgia Institute of Technology in Atlanta, has spent the past decade trying to harvest the tiny amounts of mechanical energy generated when people walk or even breathe. "We started thinking, how do we convert body motion into electricity?" he says.

His latest design uses static electricity — long thought of as a nuisance — to convert the movement of inhaling and exhaling into enough energy to power a pacemaker[10]. The generator uses two different polymer surfaces, sandwiched between electrodes and connected in a circuit. When the user breathes in and out, the surfaces touch and separate, swapping electrons — the same thing that happens when a balloon is stroked with a wool cloth. The build-up of charge causes current to flow through the wire. "Inhale and exhale, move back and forth or drive up and down and you generate power," says Wang.

Starting in 2014, Wang began testing the system in rats, creating milliwatts of energy from a device the thickness of a few sheets of paper. Now his team is testing the same technology in pigs.

Rogers' team has created[11] a biodegradable battery using electrodes made of magnesium and other metals that are safe in low concentrations and that slowly dissolve in the body. "Some devices you want to last the life of the patient. In others, you only need and want the device to be temporary," says Rogers.

PERSONAL PRIVACY

The technology could be revolutionary, but the vision of a wired-up body that sends data to an outside computer or medical centre faces a threat that already troubles the wearables industry: hacking. "When a semiconductor chip is introduced inside the body, hacking is a truly serious issue," says Someya.

One solution is to analyse data on the device itself, reducing the amount that gets sent over the airwaves. Another is to avoid the airwaves altogether. In as-yet-unpublished work, the Swedish team has developed an in-body intranet that transmits signals at low frequency using the body's water as its wires. To send information between devices, or from a device to a smartphone, users must physically touch the objects with their hands. This keeps the signals low-power and private, and avoids clogging up the data-transmitting frequencies that are already squabbled over by mobile phones and wireless routers. "It's only transmitted and exposed within your body," adds Berggren, who says that the system can already exchange data between electronically labelled objects through the body to a smartphone, and will soon integrate on-skin sensors.

However good the devices, pioneers of new materials will also struggle against a tide of medical regulation, says Malliaras. That, along with the concerns of chemical suppliers who are afraid that failing devices could leave them vulnerable to lawsuits, "puts a big brake on the adoption of new materials", he says.

Berggren and his collaborators at Acreo are among the first to try to connect a range of devices by wiring up humans. But they readily acknowledge that making the vision a reality will require multiple companies and research teams, as well as the involvement of insurance companies and health-care providers.

Berggren knows that there are big hurdles. "The challenge is to put everything together," he says. "But they did it for the car industry and it's impressive. You rarely see cars standing along the side of the road waiting for repair. Whether it's possible to do this also for humans is still a question mark, but it's definitely worth trying."

Malliaras agrees. "A car you usually keep for less than ten years," he says. "A body you want to keep for 80 or 90 years; it's a lot more precious." ■

Elizabeth Gibney *is a reporter for* Nature *in London.*

1. Kim, D.-H. *et al. Science* **333,** 838–843 (2011).
2. Kaltenbrunner, M. *et al. Nature* **499,** 458–463 (2013).
3. Schwartz, G. *et al. Nature Commun.* **4,** 1859 (2012).
4. Zhang, J. *et al. Nature Nanotechnol.* **8,** 959–968 (2013).
5. Iverson, N. M. *et al. Nature Nanotechnol.* **8,** 873–880 (2013).
6. Farra, R. *et al. Sci. Transl. Med.* **4,** 122ra21 (2012).
7. Khodagholy, D. *et al. Nature Neurosci.* **18,** 310–315 (2015).
8. Williamson, A. *et al. Adv. Mater.* **27,** 3138–3144 (2015).
9. Jonsson, A. *et al. Sci. Adv.* **1,** e1500039 (2015).
10. Zheng, Q. *et al. Adv. Mater.* **26,** 5851–5856 (2014).
11. Yin, L. *et al. Adv. Mater.* **26,** 3879–3884 (2014).

COMMENT

WEN-WEI LI

A treatment plant in Chongqing, China, which processes 40,000 cubic metres of wastewater per day.

Reuse water pollutants

Extracting carbon, nitrogen and phosphorus from wastewater could generate resources and save energy, say **Wen-Wei Li**, **Han-Qing Yu** and **Bruce E. Rittmann**.

Treating domestic and industrial wastewater so that it can be reused for drinking, irrigation and manufacturing is costly. The treatment of used household water from cooking, washing, cleaning and sanitation alone accounts for 3% of global electricity consumption and 5% of global non-carbon dioxide greenhouse-gas emissions (mainly methane). Industrial wastewater is more expensive to clean. Those proportions will rise in the next decade as the world's population grows and stricter water-quality standards are enforced by developing countries[1–3].

The costs could be more than recouped if valuable chemicals — including useful forms of carbon, nitrogen and phosphorus — were captured from wastewater. Water-treatment plants that harness methane could produce electricity rather than consume it[4], for instance. Scaled up, emerging technologies could efficiently and cheaply recover phosphate and ammonium for fertilizer.

What stands in the way of creating 'wastewater-resource factories'? Uncertainty[5,6] — about which techniques are most useful and how to combine them. Here, we outline one possible strategy for domestic water (see 'Wastewater works'), illustrating how treatment plants that now cost millions of dollars a year to run could be retuned to generate more than US$1 million a year for communities. Similar schemes applied to more diverse industrial wastewater would deliver further benefits.

DOWN THE DRAIN

Domestic wastewater contains the detritus of our daily lives — faeces, fat, food scraps, detergents and pharmaceuticals. In chemical terms, 1 cubic metre of domestic wastewater contains 300–600 grams of carbon-rich organic matter (known as carbonaceous chemical oxygen demand, or COD), 40–60 grams of nitrogen (in the form of ammonium and organic compounds), 5–20 grams of phosphorus (in phosphates and organic compounds), 10–20 grams of sulfur (mainly as sulfate) and traces of heavy metal ions.

For the past century, the bulk of domestic wastewater has been treated using the aerobic 'activated-sludge process': it is whisked with air and bacteria to oxidize the pollutants. The process is simple and is effective at removing organic compounds, nitrogen and phosphorus[7]. But it has a large energy and carbon footprint. A medium-sized plant (one that processes 100,000 cubic metres of water per day) consumes as much electricity as a Chinese town of 5,000 people (around 0.6 kilowatt-hours per cubic metre of wastewater) and emits as much CO_2 as 6,000 cars per day.

The energy embodied in the wastewater's organic matter is squandered. Also discarded are forms of nitrogen and phosphorus that would be valuable for making fertilizers. Precipitated by adding calcium, iron or aluminum salts, 90% of the phosphorus ends up buried in landfill because the precipitates cannot be taken up by plants and are often contaminated with toxic metals[8,9]. Likewise, more than 80% of the nitrogen is lost through conversion to nitrogen gas by microbes. The process also produces a lot of 'wet sludge' (5–10 kilograms ▶

per cubic metre of treated water). The drying and disposal (on land or in landfill) or incineration of this accounts for 30–50% of a treatment facility's overall costs.

Some wastewater plants digest the sludge anaerobically. Here, microorganisms in the absence of oxygen break down complex organic matter into simpler organic molecules[9], which are then converted into methane. By combusting the methane to produce electricity and heat[4], anaerobic digestion can offset 20–30% of the energy and greenhouse-gas costs of the activated-sludge process. But digestion is slow, taking 10–20 days.

PROMISING SYSTEMS

Applying anaerobic practices directly to domestic wastewater could reverse those costs entirely and generate an excess of energy, but it is not currently possible at ambient temperatures and with low concentrations of organics[9]. That could change with two new technologies being trialled — if they can be scaled up[4].

The first technology is the anaerobic membrane bioreactor (AnMBR). It uses a porous membrane to retain and concentrate solids (including particulate organic matter and the slow-growing microbes that produce methane gas) and more than 90% of the dissolved organic matter in wastewater[4]. By prolonging the materials' degradation time, it allows 25–100% more methane to be produced per cubic metre of treated water. More than 90% of the dissolved methane (at concentrations of 10–20 milligrams per litre) can be extracted with gas or vacuum techniques, using relatively little energy (less than 0.05 kilowatt-hours per cubic metre; kWh m^{-3}).

Several pilot AnMBRs have been successfully used for domestic wastewater treatment; a facility that can process 12 cubic metres per day at the Bucheon wastewater-treatment plant in South Korea has run for more than 2 years. The biggest challenge in scaling up this technology is preventing the membrane from becoming clogged, or 'fouled'. Using gas bubbles or fluidized granular activated carbon to scour the membrane surface clean requires a further 0.2–0.6 kWh m^{-3} of energy, comparable to that used in the activated-sludge process.

A second option involves microbial electrochemical cells (MXCs) that either generate electrical power directly, in the mode of microbial fuel cells, or produce energy-rich chemicals such as hydrogen gas in microbial electrolysis cells[10]. MXCs take advantage of the ability of some bacteria that — as they metabolize organic matter — transfer electrons through their cell membranes to receptors outside. If passed to the anode of a fuel cell, the electrons can deliver a current.

The products of MXCs — electricity or hydrogen gas — are more valuable and readily used than methane. But the reactions involved are slow (taking several days), notably the initial break-up of particulates, which account for half of the organic matter (COD) in domestic wastewater. A promising possibility is integrating MXCs with an AnMBR to speed up the conversion of organic matter while producing methane and electricity or hydrogen[10].

But current MXCs perform poorly on large scales. Enlarging or stacking multiple cells increases their resistance and lowers the efficiency at which energy may be recovered. Several pilot, cubic-metre-scale facilities for domestic wastewater treatment have been reported, including: one using 120-litre microbial-electrolysis-cell cassettes, installed in Howdon, UK, that recovers less than half of the electrical energy input as hydrogen gas; and a 250-litre microbial-fuel-cell unit installed in Harbin, China, that converts only 7% of the embodied energy in organic substances to electricity.

NUTRIENT RECOVERY

What of nitrogen and phosphorus? Anaerobic treatment releases them into the effluent as ammonium and phosphate ions. The effluent can be used to irrigate nearby fields. But more valuable are nitrogen and phosphorus in forms that can be stored and transported. One option is recovering both as struvite, a slow-release fertilizer that is precipitated by adding magnesium and lime. This is commercially viable at the high phosphate and ammonium concentrations (hundreds of milligrams per litre) found in sludge or livestock wastewater, but it is ineffective for domestic wastewater[8].

Two emerging technologies — ion exchange and electrodialysis — capture and concentrate phosphorus and nitrogen enough to be recovered from effluent as struvite[8]. In the first, phosphate ions are swapped with anions (such as carbonate) or ammonium ions swapped with cations (such as sodium ions) and adsorbed by materials such as iron-based hydroxides, zeolites and polymers. In the second, an electric field and membrane separate phosphorus and nitrogen ions from others on the basis of charge and size.

Both technologies are still being debugged on small scales. Problems include incomplete recovery of ions from the exchanger; the exchanger or membrane becoming blocked by organic matter; salts contaminating the concentrate; and cost. For example, membranes currently cost hundreds of dollars per square metre. And electrodialytic extraction (at a recovery rate of 90%) of phosphorus and nitrogen consumes roughly 0.23 kWh m^{-3} and 0.14 kWh m^{-3}, respectively — around two-thirds of the energy consumed in the activated-sludge process[8]. Use of MXCs may partly offset that energy input by generating electricity, but microorganisms and biomolecules aggravate membrane fouling[10].

"Nitrogen recovery from wastewater in particular would have a global impact."

Nitrogen recovery from wastewater in particular would have a global impact. In the lab, extraction of nitrogen has received less attention than has phosphorus extraction, because atmospheric nitrogen gas can be easily reduced to synthesize nitrogen fertilizer. But the process involved — the nitrogen-fixing Haber–Bosch process — is energy intensive: it accounts for a few per cent of the world's annual energy use. Substituting just 5% of the existing nitrogen-fertilizer production would save more than 50 terawatt-hours of energy, or 1.5% of China's annual electricity consumption.

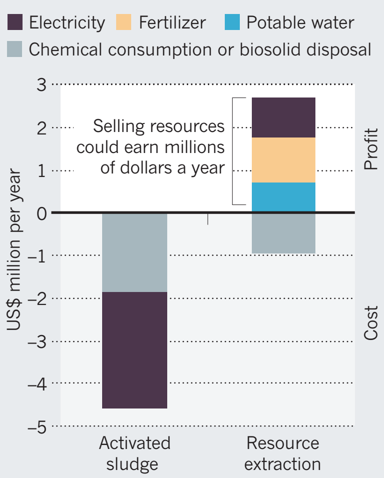
Biosolids — biomass from microbial growth and undigested faeces, fibres and other solids from the wastewater — are other by-products of anaerobic digestion that contain nitrogen and phosphorus. If they are stabilized (to avoid generating methane gas or odours) and detoxified (no pathogens or hazardous chemicals) during anaerobic treatment, they can be applied directly to the soil[5]. The United States spreads 55% of its treated biosolids onto the land, but this practice is under public and regulatory pressure because the waste is difficult to stabilize and detoxify completely, and heavy metals accumulate.

Heat treatment makes biosolids easier and safer to use. It kills pathogens, improves nutrient retention and lessens heavy-metal release. Heat from combusted methane can be used to lower energy needs[4], but the safety of biosolid products still needs to be improved and evaluated at larger scales.

The final product — water — has huge

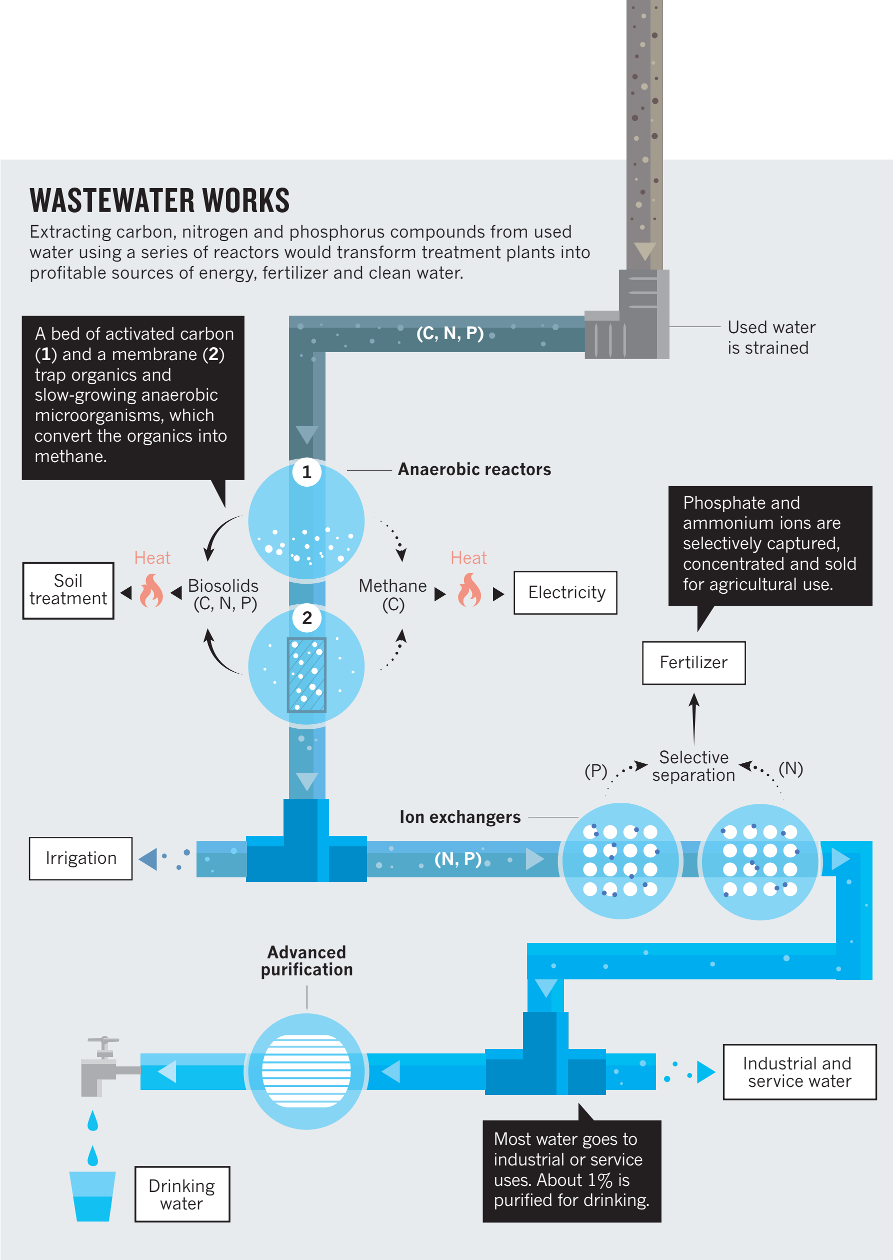
POLLUTANTS TO PROFITS

Capturing energy, nitrogen, phosphorus and water can turn wastewater treatment from a major cost into a source of profit.



Estimates for a plant processing 100,000 m^3 of wastewater per day.

SOURCE: W.-W.L., H.-Q.Y., B.E.R.



WASTEWATER WORKS

Extracting carbon, nitrogen and phosphorus compounds from used water using a series of reactors would transform treatment plants into profitable sources of energy, fertilizer and clean water.

SOURCE: W.-W.L., H.-Q.Y., B.E.R.

economic value: the global average price for potable water is $2 per cubic metre. Each type of use requires water of a different quality — from the cleanest for drinking to lower-quality water for cooling or industry uses. The treatment technology needed varies accordingly. In China, only 15% of treated water is reused and up to 98% of potable water goes to municipal and industrial sectors that could make do with lower-quality water. A 'fit-for-purpose' treatment and reuse strategy is needed.

ECONOMIC BENEFITS

We estimate that a domestic wastewater-resource factory serving a city of about half a million people in China would treat around 100,000 cubic metres of domestic wastewater per day. We calculate that each day it could produce around 17,000 kWh of electrical energy, recover 1 tonne of phosphorus and 5 tonnes of nitrogen, and reclaim 1,000 cubic metres of potable water. By contrast, an activated-sludge plant (with anaerobic digestion) of the same size would consume 50,000 kWh of electrical energy and recover no phosphorus or nitrogen. A resource factory would thus save 67,000 kWh per day (and that is without considering the energy saved in fertilizer production). This is equivalent to 1.5% of the city's daily electricity consumption.

We estimate that such a factory could yield a profit of $1.8 million per year (excluding construction costs), compared with a cost of $4.6 million per year for an activated-sludge-treatment plant (see 'Pollutants to profits'). That assumes the sale of only the 1% of water made drinkable; profits could be ten times higher if non-potable water were sold.

The economic boon could be higher still for industrial wastewaters in the agricultural, food and petrochemical sectors[1]. For example, AnMBRs can remove up to 98% of the organic matter (around 18 kilograms per cubic metre) from petrochemical effluent, producing 100 times more methane than is achievable with domestic wastewater. Livestock wastewater is rich in organic molecules and phosphorus, making it an important potential source of energy and fertilizer[8].

Government support will be crucial to developing wastewater-resource factories and promoting a sustainable water-resource market. For the next decade, extracting resources from wastewater will remain expensive relative to fossil-fuel energy and current processing methods. Why? Because environmental costs are not yet factored into pricing and emerging recovery technologies have not yet benefited from economies of scale. Priorities will change as energy, resource and global-warming stresses intensify.

What next? Governments must establish regulatory frameworks that include the costs of waste disposal and greenhouse-gas emissions. They must invest in demonstrations at scale of the pre-commercial or early-adopter technologies; initially subsidize the sales of recovered products; and promote the benefits of the recycled-resource concept.

Governments and enterprises in the sector should provide targeted research funds as well as land and infrastructure. To ensure that the products are suitable, technological development must involve input from regulators, managers of wastewater facilities, engineers, researchers and the public.

National initiatives are needed that suit local environmental, economic and social conditions. Industrialized countries should integrate the emerging processes when they replace ageing treatment facilities. And emerging economies such as China and India should incorporate them as they expand their water-treatment capacities. ■

Wen-Wei Li *is associate professor and* **Han-Qing Yu** *is professor of wastewater systems and sustainability at the Chinese Academy of Sciences' Key Laboratory of Urban Pollutant Conversion, University of Science & Technology of China, Hefei, China.* **Bruce E. Rittmann** *is professor of environmental engineering and director of the Swette Center for Environmental Biotechnology, Arizona State University, Tempe, Arizona, USA.*
e-mails: hqyu@ustc.edu.cn; rittmann@asu.edu

1. Heidrich, E. S., Curtis, T. P. & Dolfing, J. *Environ. Sci. Technol.* **45,** 827–832 (2011).
2. Koppelaar, R. H. E. M. & Weikard, H. P. *Glob. Environ. Chang.* **23,** 1454–1466 (2013).
3. US Environmental Protection Agency. *Global Anthropogenic Non-CO_2 Greenhouse Gas Emissions: 1990–2030* (2012).
4. McCarty, P. L., Bae, J. & Kim, J. *Environ. Sci. Technol.* **45,** 7100–7106 (2011).
5. Peccia, J. & Westerhoff, P. *Environ. Sci. Technol.* **49,** 8271–8276 (2015).
6. Verstraete, W., Van de Caveye, P. & Diamantis, V. *Bioresour. Technol.* **100,** 5537–5545 (2009).
7. Rittmann, B. E. & McCarty, P. L. *Environmental Biotechnology: Principles and Applications* (McGraw-Hill, 2001).
8. Rittmann, B. E., Mayer, B., Westerhoff, P. & Edwards, M. *Chemosphere* **84,** 846–853 (2011).
9. Smith, A. L. *et al. Environ. Sci. Technol.* **48,** 5972–5981 (2014).
10. Li, W. W., Yu, H. Q. & He, Z. *Energ. Environ. Sci.* **7,** 911–924 (2014).

Most soils are in private ownership, making it tricky to implement binding international agreements.

Govern our soils

Luca Montanarella calls for a voluntary international agreement to protect the ground beneath our feet from erosion and degradation.

Eighty years ago, in 1935, soils were for the first time officially recognized as a limited national resource that should be responsibly managed. In the wake of the catastrophic erosion that caused the infamous Dust Bowl drought, the US government passed the Soil Conservation Act. "The history of every Nation is eventually written in the way in which it cares for its soil," wrote President Franklin D. Roosevelt.

Roosevelt's act was largely successful. It encouraged farmers to apply sustainable management practices — such as tilling less, installing windbreaks, and planting along slope contours[1]. Between 1982 and 2007, soil erosion in US cropland declined by 43% (ref. 2).

The history now being written in the world's soils is not so rosy. Every year, 75 billion tonnes of crop soil are lost worldwide to erosion by wind and water, and through agriculture; this costs about US$400 billion a year[3]. Only a few countries have national legislation protecting soil, including Germany and Switzerland[4]. Attempts at binding international legal agreements have so far failed.

This cannot go on. Soils are a limited natural resource, unequally divided between nations and people. They provide fertilizer for growing food; store and filter water; host rich ecosystems, including many little-known species; provide resources such as peat, sand, clay and gravel; and hold our cultural and historical memory in archaeological artefacts. The ground beneath our feet is a public good and service.

GET OFF MY LAND

Without governance to assure wise management and equitable access, we are heading towards increased poverty, hunger, conflict, land grabs and mass migration of displaced populations, such as that seen during the Great Depression[5]. The world now stands at a moment of opportunity. A Global Soil Partnership (GSP) exists, and could implement a voluntary system of global governance. But the GSP needs to develop clear, concrete proposals for action to secure more funding and move forwards.

International soil governance faces great challenges. Take, for example, a nearly decade-long attempt by the European Union to implement a governance framework. A team at the European Commission (of which I was part) developed a common EU strategy for soil protection[6] including a proposed EU Soil Framework Directive, which would have obliged member states to take action to prevent soil degradation. It was the result of several years of consultations in specialized working groups that included scientists, policymakers, industry representatives, landowners and farmers, as well as concerned non-governmental organizations (NGOs) and other stakeholders. Much was at stake, including the ongoing, costly remediation of more than 3 million contaminated sites in Europe, such as old industrial areas and mining sites, and the question of who should pay.

Several EU member states opposed the directive. Their arguments were much the same as those used in 1935 by opponents to the US Soil Conservation Act. They countered that soils are a strictly local issue, and should be governed locally rather than

by a central authority (the subsidiarity principle). They noted that because most soils are privately owned, they should not fall under the remit of public governance, and pointed out that soils do not move, and therefore there is no need for transnational or global governance instruments. After some debate and a long period of apathy, the directive was withdrawn by the European Commission in May 2014.

The counter-argument is simply that good-quality soils are necessary for the food, fibre and fuel of a growing population. That makes soil — like air and water — a shared resource that requires governance. And, because most soils are indeed privately held, legally binding international agreements are unrealistic. Instead, governance must be based on voluntary efforts by national governments, local land owners and administrations.

Progress so far has been disappointing. In 1982, the Food and Agriculture Organization of the United Nations (FAO) adopted a World Soil Charter with 13 recommendations for sustainable soil management. It enshrines some basic principles such as: "the use of these resources should not cause their degradation or destruction because man's existence depends on their continued productivity". That charter was endorsed by all members of FAO (nearly all national governments). It remains largely ignored.

"Soils are necessary for the food, fibre and fuel of a growing population."

The dramatic rise in food prices during the 2008 global food-commodities crisis finally raised the attention of policymakers. That led to the creation in 2011 of the FAO's GSP: a voluntary body tasked with finally enacting the soil charter's principles.

TIME FOR LEGISLATION

The GSP has concentrated its activities on promoting sustainable management of soils, for example by encouraging consistent research, education and good policy. In 2016, it will launch a World Soil Prize to reward best practice. Concrete action on the ground is in the hands of Regional Soil Partnerships that include all local stakeholders. So far, most of the GSP's work has been in organizing conferences and developing task-force plans of action. Sadly, these mostly provide vague expressions of intent. Four years after its creation, the GSP is under increasing pressure from NGOs and funders to deliver results.

The GSP's clearest call is for the development of a Global Soil Information System. Unfortunately, the GSP failed to establish a comprehensive partnership with everyone involved, and as a result several parallel independent projects have emerged, such as the GlobalSoilMap.net consortium and the Global Soil Information Facilities. Bringing all of these efforts together will be difficult.

The 1930s Dust Bowl drought prompted the first soil-conservation act, in the United States.

To underpin the GSP, an Intergovernmental Technical Panel on Soils (ITPS; of which I am chair) was established in June 2013. Like the Intergovernmental Panel on Climate Change, the ITPS aims to provide scientific and technical guidance to policymakers. It is composed of 27 soil experts from across the seven FAO regions. Our ambition is to serve the GSP and all soil-related multilateral environmental bodies, such as the United Nations Convention to Combat Desertification, the Convention on Biological Diversity and the United Nations Framework Convention on Climate Change.

The main product of the ITPS's first two years is the Status of World's Soil Resources report, scheduled for release at the closing ceremony of the UN International Year of Soils in December 2015. The report, the first comprehensive assessment of global soil resources, is the collaborative effort of more than 200 scientists. It highlights serious concerns such as nutrient imbalance: some parts of the world suffer from an excess of fertilizer use, whereas much of the developing world suffers from a severe lack of fertilizers. The ITPS is preparing practical recommendations for reversing these trends.

The GSP is the best current option for driving forward those recommendations, despite its shortcomings. The partnership needs to motivate all invested parties to develop commitments to specific actions. These should enshrine soil management in legislation tailored to each country's needs. The GSP needs to prove that it can be more than just a talking shop, and can generate political will and raise funding. The FAO has suggested an initial budget of $64 million over five years for the GSP[7], mainly to help to develop the Global Soil Information System and to promote training and capacity building in developing countries. So far, less than 10% of that has been raised from donors, mainly the European Commission.

Increasingly, people speak of 'soil security'[8], in analogy with food and water security. In a world facing increasing stress from a growing, hungry population and changing climate, soils will become ever more important. ■

Luca Montanarella *is a senior expert at the Joint Research Centre of the European Commission in Ispra, Italy, and chair of the Intergovernmental Technical Panel on Soils. e-mail: luca.montanarella@jrc.ec.europa.eu*

1. Agrabright, M. S. *et al. Historical Changes in Soil Erosion 1930–1992* (NRCS, 1996); available at http://go.nature.com/6vc4oc
2. *2007 National Resources Inventory* (NRCS, 2010); available at http://go.nature.com/uhj92u
3. Noel, S. *et al. Reaping Economic and Environmental Benefits from Sustainable Land Management* (ELD Initiative, 2015); available at http://go.nature.com/aft1f3
4. Montanarella, L., Arnold, R. & Micheli, E. in *Advances of GeoEcology* 40 (Catena, 2009).
5. van Schaik, L. & Dinnissen, R. *Terra Incognita: Land Degradation as Underestimated Threat Amplifier* (Netherlands Institute of International Relations Clingendael, 2014).
6. Commission of the European Communities. *Thematic Strategy for Soil Protection* (European Commission, 2006).
7. Global Soil Partnership. *Healthy Soils Facility of the Global Soil Partnership, Programme Document* (FAO, 2013).
8. McBratney, A., Field, D. J., & Koch, A. *Geoderma* **213**, 203–213 (2014).

CHRISTIAN ALS/PANOS PICTURES

Port-au-Prince after the devastating 2010 earthquake that killed more than 85,000 Haitians.

GEOPHYSICS

Vast forces underfoot

Andrew Robinson examines three books that see seismicity as both grimly destructive and, in some contexts, culturally energizing.

A blink in geological time — 150 years — has passed since Jules Verne published his fantasy *Journey to the Centre of the Earth* (see D. Chatelain and G. Slusser *Nature* **513,** 169–170; 2014). Half a century later, geologist and meteorologist Alfred Wegener published his radical theory of continental drift, *The Origin of Continents and Oceans* (see T. Nield *Nature* **526,** 192–193; 2015). And half a century after that, in 1965, the theory of plate tectonics — partly inspired by Wegener — was established by geophysicist John Tuzo Wilson among others.

In the subsequent half-century, exploration of the Solar System has revealed that Earth is the only planet in it with a global system of plate tectonics. Satellites in the US Global Positioning System monitor plate movements with an accuracy of a few millimetres. But our understanding of Earth's mantle and core is much less advanced. The deepest borehole penetrates just 12,262 metres, two-thousandths of Earth's radius. Reliant mainly on seismographic monitoring, modelling and post-quake analysis, geophysicists and seismologists remain perplexed about the exact structure of the inner core and the precise cause of earthquakes. Developments in seismology since the 1930s, when Charles Richter invented his local-magnitude scale, have ranged from laboratory fault-friction experiments to global seismic tomography. But our ability to predict the timing, location and magnitude of earthquakes has scarcely progressed.

Journey to the Centre of the Earth: The Remarkable Voyage of Scientific Discovery into the Heart of Our World
DAVID WHITEHOUSE
Weidenfeld & Nicolson: 2015.

Earthquake Time Bombs
ROBERT YEATS
Cambridge University Press: 2015.

Impact of Tectonic Activity on Ancient Civilizations: Recurrent Shakeups, Tenacity, Resilience, and Change
ERIC R. FORCE
Lexington: 2015.

Now, three books examine earthquakes from distinct angles. In his *Journey to the Centre of the Earth*, astronomer and BBC science broadcaster David Whitehouse takes the reader on a scientific journey from crust to core in a book inspired by Verne's, but making slight reference to it. Seismologist Robert Yeats, in *Earthquake Time Bombs*, focuses on the crust, and how to protect vulnerable conurbations — his "time bombs" — from probable seismic shocks. And geologist ▶

A first-century-AD relief in Pompeii, Italy, shows the Temple of Jupiter tilting after an earthquake in AD 62 or 63.

DEA/A. DAGLI ORTI/DE AGOSTINI/GETTY

and geoarchaeologist Eric Force investigates earthquakes from the third millennium BC in *Impact of Tectonic Activity on Ancient Civilizations*, theorizing that they stimulated trade and helped to shape civilizations.

Whitehouse's account is the most readable and wide-ranging, although it is inevitably speculative. "We will reach the distant stars before we reach the centre of the Earth," he writes, after descending more than 1,000 metres into one of the deepest mines in Europe, the Boulby potash mine in northeast England. He is also the most adept at mixing the history of Earth science — beginning with Edmond Halley's maritime expedition to measure Earth's magnetic field around 1700 — with comments by current researchers. One of them admits that "everything" about the inner core — structure, anisotropy, topography and dynamics — "is getting increasingly complex as we get more data".

However appealing, Whitehouse's account contains errors. For instance, the disastrous 1906 San Francisco earthquake, in which more than 3,000 people died, did not prompt the introduction of "building and emergency regulations"; those came decades later in California. Indeed, San Franciscans did their best to blame the city's destruction on the fire started by the earthquake and to carry on with 'business as usual'. Nor did seismologist Beno Gutenberg, mentor of Richter, flee persecution in Germany in 1933 for a job at the University of California. He left in 1930, before Adolf Hitler came to power, and settled at the California Institute of Technology in Pasadena, along with a visiting Albert Einstein.

Yeats's book is a follow-up to his magnum opus, the specialist *Active Faults of the World* (Cambridge University Press, 2012). *Earthquake Time Bombs* aims to reach a wider audience. Writing that the "next great earthquake will be a disaster, but failing to prepare for it will lead to a catastrophe", he recounts how in early 2010 he told a *Scientific American* reporter that Port-au-Prince should be regarded as a time bomb. Its swelling population occupied dilapidated slums adjacent to a plate-boundary fault that had not sustained a major earthquake since the mid-eighteenth century. A week after he made his comments, a magnitude-7 earthquake destroyed the Haitian capital, killing at least 85,000 people (the government put the figure at more than 300,000), mainly as a result of inadequate and corrupt building practices (R. Bilham *Nature* **502,** 438–439; 2013). Yeats had not, of course, predicted the quake. He was simply aware of research on the fault published in 2008 by Eric Calais and his colleagues (D. M. Manaker *et al. Geophys. J. Int.* **174,** 889–903; 2008) These researchers had privately alerted the Haitian government, but advised that they could not predict the timing of the recurrence.

Sixty of the world's largest cities lie on plate boundaries and are at risk from interplate earthquakes. Yeats duly discusses the usual suspects, such as San Francisco, Tokyo, Istanbul and Santiago. But he also explores less familiar threats, including the Cascadia subduction zone (Seattle, Portland and Vancouver), where he lives, along with Tehran, Kabul, parts of the Himalayan region, Manila, Caracas, Wellington and the East African Rift Valley. Surprisingly, he neglects the hazard from intraplate quakes that occur away from boundaries, including that in Gujarat, India, in 2001 and, most famously, the 1811–12 earthquakes in Missouri in the middle of the North American plate. The Missouri quakes have provoked much debate among leading US seismologists such as Susan Hough and Seth Stein, author of *Disaster Deferred* (Columbia University Press, 2010), a book that Yeats does not mention. He concludes with convincing grimness that only California, Japan, Chile and New Zealand have taken the earthquake hazard seriously.

NATURE.COM
For more on science in culture see:
nature.com/booksandarts

Where Yeats expounds on the destructive power of quakes, Force posits that they may have rocked the cradles of past civilizations. High tectonic activity has accompanied the birth and growth of many ancient civilizations in the Middle East, Greece and Italy and, to a lesser extent, the Indus Valley and China. During the second and first millennia BC around the Mediterranean Sea, the Minoan, Mycenaean, Greek, Etruscan and Roman civilizations arose during eras of major seismic activity in their regions. No comparable cultures developed on the relatively inactive coasts of Spain, France and Libya, observes Force. He suggests that frequent tectonic activity was a "long-term cultural stimulant", forging ancient communities that were resilient, cooperative, innovative and outgoing, and where "elders would be passing on an expectation of change to younger generations".

It is a tantalizing thesis, which Force pursues tenaciously and with considerable skill. However, the book perhaps goes too far in its claim for the dominance of seismic activity in the development of civilizations. If the hypothesis is correct, how did Egypt, which had (and has) relatively low tectonic activity, produce a major civilization? Surely, climate, coasts, rivers, fertile soil and supplies of water, minerals, building materials and fuel are also key, even if some of these factors are also influenced by tectonic activity.

Nevertheless, Force's speculation remains an intriguing possibility. Today, both Silicon Valley and Hollywood lie on the San Andreas fault. Just a coincidence? Or is the hidden cause of these powerhouses of imagination and innovation that the region has frequently been "all shook up"? ■

Andrew Robinson *is the author of three books on earthquakes, including the forthcoming* Earth-Shattering Events: Earthquakes, Nations and Civilization.
e-mail: andrew.robinson33@virgin.net

PSYCHOLOGY

The scarred self

Anthony King reviews an exhibition on the horror, and hope, posed by trauma.

Less than a month after the terrorist attacks in Paris, Trauma, the latest show at Science Gallery Dublin, feels unnervingly relevant. Exploring, among other things, the surprising role of trauma in emotional resilience, this collection of objects and ideas ranges from haunting photos from Northern Ireland's Troubles to a room-sized instrument crafted by a composer with tinnitus (see J. Hoffman *Nature* **505,** 159; 2014) and photograms of plant specimens from Chernobyl.

Trauma, from the Greek word meaning 'wound', is profoundly personal: the sufferer is sealed off within the mental or physical experience. Katharine Dowson's *Memory of a Brain Malformation*, a delicate laser etching of a brain tumour in glass, emphasizes this isolation. Dowson — who often works with scientists and physicians — portrays the growth as a discrete entity inside a nest of sinuous veins. (The actual tumour was successfully removed from her cousin's brain by laser treatment.) The work evokes both the emotional trauma of diagnosis and the energy of a positive outcome.

External trauma to the head can be just as damaging, and the neurological problems arising from it in sport are the focus of intense research. The installation *Impact* examines the design of helmets for sports such as American football and Irish hurling, sparked by such studies. Mechanical engineer Ciaran Simms at Trinity College Dublin, for instance, examines the body's response to high-force impacts in rugby; Stefan Duma at the Virginia Polytechnic Institute and State University in Blacksburg uses real-time sensors in the field and lab to rate commercial helmet design.

Trauma: Built to Break
SCIENCE GALLERY DUBLIN
Until 21 Feb 2016.

A darker realm of trauma is explored on the gallery's ground floor in *The Interrogation of Detainee 063*, an infographic detailing 50 harrowing days in the interrogation of Mohammed al-Qahtani at the United States' Guantanamo Bay camp in Cuba. The exhibit underlines the extreme suffering triggered by torture. Colour-coding shows the duration of interrogation, loud music and inhumane and degrading treatments, such as being forced to wear a muffling hood or humiliating signs, or to write letters of apology to victims of the terror attacks of 11 September 2001.

Upstairs, *Stressed Body, Stressed Brain* investigates one physiological response that is central to the notorious torture technique waterboarding. This response, the diving reflex, is triggered when the face is immersed in cold water. The exhibit invites viewers to lie down and have a damp cloth placed on their cheeks to gauge how this upsets memory recall and slows heart rate by as much as one-quarter. Its curators are physiologist Áine Kelly and neuroscientist Shane O'Mara, author of *Why Torture Doesn't Work* (Harvard University Press, 2015; see L. T. Harris *Nature* **527,** 35–36; 2015).

O'Mara has shown elsewhere how stress and trauma can trigger the creation of false memories. *Memory Laundering* — essentially a large cabinet holding dozens of deposit boxes — plays with this mutability. Created by makers Designgoat, It is inspired by the work of neuroscientist Susumu Tonegawa and the team at the RIKEN–MIT Center for Neural Circuit Genetics at the Massachusetts Institute of Technology in Cambridge, who collaborated with the gallery. You are asked to write down one good and one bad memory, and place them in one of the boxes. When you return to retrieve them, the details have been edited by a gallery mediator concealed behind the cabinet.

The most graphic of the exhibits is a series of photographs of an operating theatre in Afghanistan's Helmand Province. *Sightlines I/Supernumerary* by installation artist David Cotterell is a record of his stint as an embedded photographer with the UK Joint Forces Medical Group. During it, he created diptychs and triptychs of the visceral business of emergency medicine — containing and controlling trauma. The images reference the dramatic chiaroscuro of painters such as Caravaggio, evoking horror yet suggesting sublime beauty. Alongside the bloody collage stands *XSTAT 30 Hemorrhage Control Device*, an innovative syringe containing 92 miniature cellulose sponges, designed to control severe bleeding.

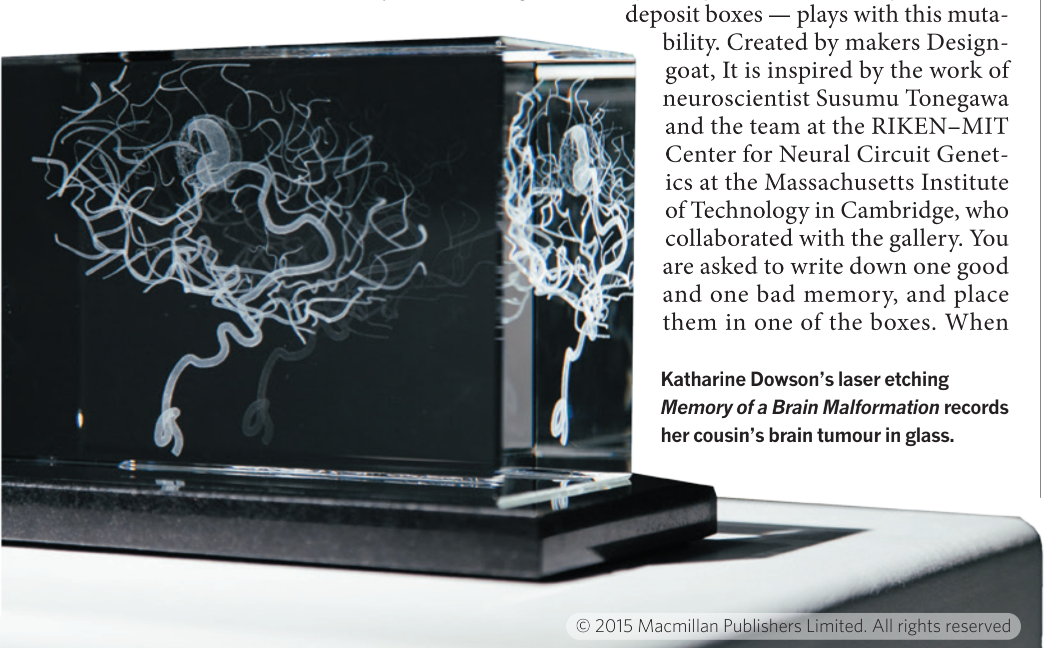
"Trauma is something that life can profit from, enhancing resilience."

"Trauma is the ultimate insult," says co-curator and neuroscientist Daniel Glaser. It is about not the moment, but the aftermath, he adds — a truth easily observable among victims from Syria to France. Dublin has had its own share of trauma. Next year sees the centenary of the Easter Rising, when Irish republicans proclaimed independence from Britain. Six locations across the city will be stamped with a bandage symbol on a map available at the gallery as part of artist Sarah Bracken's *Bandage*, marking hidden scars from pitched street battles, arrests and executions by the British military during the rising.

At base, this is a show about recovery. The tumour is excised; the blood flow is staunched; life goes on. From mental wounds to psychological damage — historically viewed as inevitable aspects of the human condition — the message in Trauma is ultimately positive. O'Mara explains why. Between 30% and 70% of traumatized people experience post-traumatic growth, he writes: their suffering opens up "new perspectives not previously available to them". As he notes, "trauma is something that life can profit from, enhancing resilience, and providing lessons to us all". ■

Anthony King *is a writer based in Dublin.*
e-mail: anthonyjking@gmail.com

SCIENCE GALLERY AT TRINITY COLLEGE DUBLIN



Katharine Dowson's laser etching *Memory of a Brain Malformation* records her cousin's brain tumour in glass.

Correspondence

It is rational to protect Antarctica

We are dismayed that the international commission that oversees the Convention on the Conservation of Antarctic Marine Living Resources has voted against establishing marine protected areas (MPAs) around Antarctica for the fifth consecutive time. These MPAs are designed to protect wildlife hotspots of world significance.

The main opponents were member states that fish or intend to fish for toothfish (*Dissostichus* spp.) and Antarctic krill (*Euphausia superba*). Toothfish, which are sold as Chilean sea bass, are the top fish predators in the Southern Ocean; krill is a crucial component of the marine food web that is sold as fishmeal and for fish-oil pills.

The convention's goal of conservation is being marred by some member states who are misinterpreting the "rational use" proviso in its text. Originally intended to allow fishing in the Southern Ocean only if it complied with strict guidelines, this term is being misinterpreted as an unrestricted right to fish and as an excuse to block tighter regulations (see J. Jacquet *et al. Mar. Policy* **63,** 28–34; 2016).

The commission operates by consensus, so a single member state can prevent cooperation. This year, China and Russia blocked the proposed MPAs for the east Antarctic — even though these included boundaries designed to accommodate fisheries — and Russia blocked an MPA in the Ross Sea.

Jennifer Jacquet *New York University, New York, USA.*
Cassandra Brooks *Stanford University, California, USA.*
jacquet@nyu.edu

Mining disaster: huge species impact

On 5 November, a huge mudflow contaminated with iron ore from mine workings was released into the Rio Doce river in southeast Brazil after two dams broke. Immediate action is necessary to evaluate the massive human and ecological impact of this catastrophe, and there must be a concerted effort to prevent further such incidents.

As well as killing several people, the accident threatens the water supply of many large cities downstream that are already severely limited by a long-standing drought. The polluted river runs through the Atlantic rainforest and is likely to damage the exceptional endemic fauna and flora in its waterways.

Of the 71 recognized fish species in the river, 11 were considered endangered before the mud slide (see go.nature.com/zmry1z; in Portuguese). The accident also interrupted reproductive migrations for many of these species.

Markus Lambertz *Zoological Research Museum Alexander Koenig, Bonn, Germany.*
Jorge A. Dergam *Federal University of Viçosa, Minas Gerais, Brazil.*
m.lambertz@zfmk.de

Mining disaster: restore habitats now

In Brazil's Atlantic rainforest region last month, cities were flooded and watersheds contaminated when some 50 million cubic metres of heavily polluted water was released from an iron-ore tailings pond. The mining company responsible and Brazil's environment ministry should act swiftly to mitigate the human and ecological damage.

The release has deprived some 500,000 people of their water supply. It is likely to damage the entire ecological network through chemical pollution, reduced oxygen availability and high turbidity, further threatening the region's status as one of the world's biodiversity hotspots.

Authorities will need to collaborate with universities on ecosystem restoration and revitalization projects.

Jhonny Capichoni Massante *Federal University Fluminense, Niterói, Rio de Janeiro, Brazil.*
jcmassante@id.uff.br

Star universities in the Muslim world

As former chairman of Pakistan's Higher Education Commission and former coordinator-general of the Organisation of Islamic Cooperation's science and technology body COMSTECH, I suggest that some universities in the Muslim world are not in such dire need of revitalization as Nidhal Guessoum and Athar Osama imply (*Nature* **526,** 634–636; 2015).

At least 3 such institutions are ranked in the world's top 250 — the University of Malaya in Kuala Lumpur, and King Fahd University and King Saud University, both in Saudi Arabia (see go.nature.com/4gfu2u). In 2013 and 2014, the Middle East Technical University, Istanbul Technical University and Bilkent University in Turkey were ranked in the top 400 globally (see go.nature.com/m6195d). Pakistan's National University of Sciences and Technology and the Pakistan Institute of Engineering and Applied Sciences were ranked in the top 200 Asian universities in 2014 (see go.nature.com/kdwt8w). The King Abdullah University of Science and Technology in Saudi Arabia and the Masdar Institute in Abu Dhabi are rising stars.

According to 2014 data on scientific publications, Iran ranks 16th in the world, Turkey is 19th and Malaysia is 23rd — on a par with Switzerland, Taiwan and some Scandinavian countries, and ahead of South Africa (see go.nature.com/ms6fct).

Furthermore, the requirements of the United Arab Emirates' Commission of Academic Accreditation (CAA) are more stringent than those of the US Accreditation Board for Engineering and Technology (ABET), for instance. Whereas the CAA requires faculty members to have the highest degree in their field (such as a PhD), ABET requires only appropriate qualifications. The CAA also requires universities to have accredited PhD programmes in addition to accredited bachelor's and master's degrees.

Javaid Laghari *Pasadena, California, USA.*
jlaghari@gmail.com

Microbiome studies need local leaders

As researchers on the Brazilian Microbiome Project, we contend that creating a robust International Microbiome Initiative (IMI) needs local leadership rather than top-down scientific unification (see N. Dubilier *et al. Nature* **526,** 631–634; 2015).

Microbial diversity and function are tied to geographically relevant features, so local investigation of these peculiarities is needed to underpin national biodiversity-protection measures. Researchers attached to such projects can boost their country's reputation in science and technology. If the IMI succumbs to pressure to avoid local research consortia, it could bias scientific priorities and project management towards the interests of a few, and compromise the independent verifiability of the science.

Resources expended on global collaborations without a clear description of aims could also result in an endless development of standards and protocols (see *Nature* http://doi.org/9gx; 2015). In our view, it is important to unite researchers locally to discuss such issues before imposing a pre-established model.

Victor S. Pylro, Daniel K. Morais *René Rachou Research Center (CPqRR-FIOCRUZ), Belo Horizonte, Minas Gerais, Brazil.*
Luiz F. W. Roesch *Federal University of Pampa, São Gabriel, Rio Grande do Sul, Brazil.*
victor.pylro@brmicrobiome.org

Lisa Jardine

(1944–2015)

Historian of science who chaired pioneering embryology regulator.

LISA JARDINE

Lisa Anne Jardine rewrote the history of European intellectual and scientific life in the sixteenth and seventeenth centuries. She was always most interested in the precise ways in which her research subjects did their jobs. She showed how the sixteenth-century theologian Desiderius Erasmus invented a new kind of career as a scholar and writer in the world of print; how Robert Hooke and Christopher Wren transformed the city of London's landscape in the seventeenth century; and how the Huyghens family and their Dutch compatriots created a sparkling world of exotic gardens, spectacular works of art and penetrating inquiries into nature.

She told these stories in huge biographies and histories that were as accessible and elegant in style as they were novel in content. These included *The Curious Life of Robert Hooke* (HarperCollins, 2003) and *Going Dutch* (Harper, 2008). Latterly, she steered the United Kingdom's pioneering regulator, the Human Fertilisation and Embryology Authority (HFEA) safely through choppy waters.

Jardine, who died of cancer on 25 October 2015, was professor of Renaissance studies at University College London. Born on 12 April 1944 in Oxford, Jardine was in some ways fated to study and write about science and the humanities. Her father, mathematician and biologist Jacob Bronowski, created the landmark 1973 BBC documentary series, *The Ascent Of Man*. Like him, Jardine read mathematics at the University of Cambridge, later switching to study English. She then did an master's in translation at the University of Essex and a doctorate in Renaissance studies at Cambridge.

She became fascinated with what might be called the history of knowledge and the practices by which humans attain it. She admired dominant historical figures of the Royal Society in London, such as Hooke and Wren, and showed how they found new ways to unpick the fabric of nature. In 1996 she devoted a pioneering book on the Renaissance, *Worldly Goods* (Macmillan), to the merchants and customers of the fifteenth and sixteenth centuries who learned to appreciate fine paintings, sumptuous fabrics and rare objects. The Renaissance itself, in her view, emerged from their finely honed consumerism.

As Jardine's interests developed, she invented new ways of writing history. Scholars knew for centuries that the sixteenth-century writer Gabriel Harvey adorned his books with vast marginal notes. Jardine deciphered them — and identified Harvey as a figure of a previously unknown kind, a Renaissance political adviser who served great men by reading the classics with them. Historians of science, in the 1980s and after, concentrated on reconstructing the precise, local practices of men such as Hooke and Robert Boyle. Jardine did the same: but she never forgot that they were polyglots, in dialogue with others across Europe. English science in its early heyday, as she portrayed it, was not a creation of national genius but a structure raised on foundations laid by the Dutch.

In the 1970s and 1980s, when women were still rare in academia, Jardine became a mentor and model for a great many younger scholars, both male and female. Her greatest talent — and greatest love — was teaching. She lectured to prodigies and ordinary students with equal engagement, mentored brilliant scholars with immense generosity, and always found a way to look after one more student than the budget allowed. Even more than her books, her students are her monument.

Committed to public service, Jardine held many high-profile posts. She judged the Man Booker and Whitbread literary prizes, and served as a trustee for London's Victoria and Albert Museum and a council member of the Royal Institution. In her two terms as chair of the HFEA, from 2008 to 2014, she led efforts to reduce multiple births resulting from *in vitro* fertilization (IVF), provide fairer compensation for donors, reduce regulatory overlap and give better access to data for researchers — all this at a time when the future of the globally respected organization was uncertain.

She felt particularly honoured to have overseen the 2012 public consultation on mitochondrial replacement. This IVF technique aims to prevent women from passing on harmful mutations in the cell's energy-producing structures, mitochondria, by using a third party to provide healthy mitochondrial DNA for a future baby. The United Kingdom this year became the first country in the world to allow this technique in the clinic, and the HFEA's engagement exercise is frequently cited in current debates on genome editing in sperm, eggs and embryos.

What mattered most to Jardine was not the institution she served, but the quality of her service. Of her many distinctions, election to the Royal Society as an honorary fellow particularly delighted her. She was just as proud of her stint as governor of a London school.

Her favourite brooch read *multum in parvo* ('a lot in a little') — her joke about her height. She was a commanding presence in public, a dazzling speaker whose plenary lectures were the most memorable events at many conferences. A wider British public knew her from many years of broadcasts, including the 2013 BBC radio series *Seven Ages of Science*, in which she vividly conveyed the excitement and complexity of historical research. Her voice on the page was distinctive: trenchant, accurate and unfailingly eloquent, whether she was arguing a historical case in a journal or engaging in a contemporary debate in a newspaper or online article.

Lisa was a rare figure — she combined academic brilliance with a deep commitment to public service, and made it all look so easy. ■

Anthony Grafton *is professor of history at Princeton University in Princeton, New Jersey, USA. He collaborated extensively with Lisa Jardine.*
e-mail: grafton@princeton.edu

NUCLEAR PHYSICS

Close encounters of the alpha kind

Breakthrough calculations of collisions between two helium nuclei pave the way to a quantitative understanding of how the elements carbon and oxygen were made in stars — and to improved models of stellar evolution. SEE LETTER P.111

SOFIA QUAGLIONI

The life-enabling elements carbon and oxygen were mainly made in red giant and supergiant stars, through a sequence of fusion reactions known as helium burning. In this process, helium nuclei (^{4}He) — aggregates of two protons and two neutrons, named alpha particles by Ernest Rutherford[1] — are progressively converted into carbon and oxygen nuclei (^{12}C and ^{16}O, respectively). But exactly how each of these reactions happens at a fundamental level remains unexplained. On page 111 of this issue, Elhatisari *et al.*[2] take a crucial step towards addressing this question by describing the collision of two alpha particles (alpha–alpha scattering) from first principles. The theoretical–computational methods described in this work could also be used to characterize collisions between certain other composite quantum particles.

Helium burning starts when two alpha particles collide with each other and attempt to fuse. But the product of that fusion is an unstable beryllium isotope, ^{8}Be, which decays almost instantly back into two helium nuclei. A third alpha particle therefore has to be captured nearly simultaneously with the collision of the original pair for ^{12}C to be formed. This process is known as the triple-alpha reaction[3] and was first proposed in 1952. Oxygen is then created when ^{12}C captures a fourth alpha particle[4] (Fig. 1). The ratio of the amount of carbon to that of oxygen produced through these processes has profound repercussions on the later evolutionary phases of magnificent massive stars such as Orion's Betelgeuse, and their ultimate fate once they explode as supernovae[5].

The odds of any of these reactions taking place are small at the low energies encountered in stellar environments, largely because the positively charged colliding nuclei electrically repel each other. The low reaction rates make helium-burning reactions difficult to replicate and measure in a laboratory. This prevents the carbon-to-oxygen ratio produced in stars from being accurately estimated, and introduces large uncertainties in stellar-evolution models and simulations of the processes that create nuclei (nucleosynthetic processes). A computational approach capable of describing helium burning from first principles could come to the rescue by providing 'measurements' from simulations. Of course, before such theoretical predictions can be trusted, their accuracy in reproducing experimentally determined data must be ascertained. The scattering of two alpha particles has been characterized experimentally, and is a good starting point in this case.

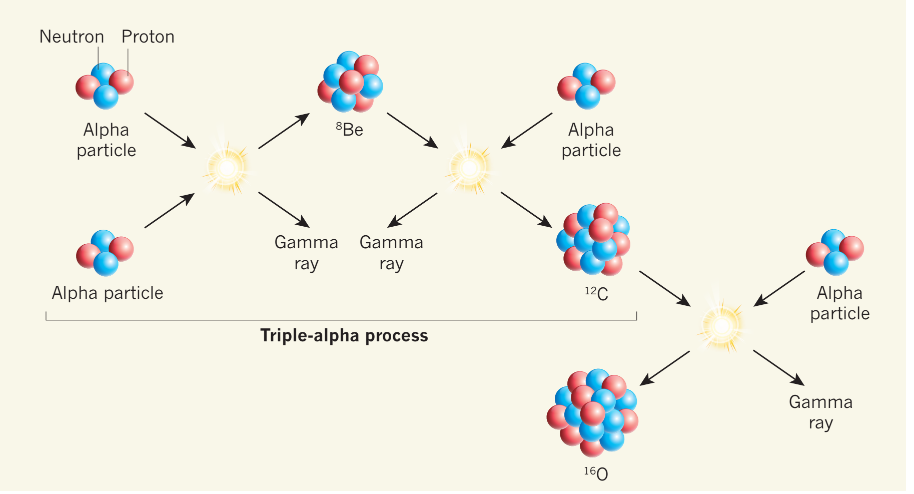


Figure 1 | Formation of carbon and oxygen nuclei. Carbon nuclei, ^{12}C, form in stars through a sequence called the triple-alpha process. Two alpha particles (which consist of two protons and two neutrons) collide to produce a transient beryllium nucleus, ^{8}Be, and a gamma ray. If another alpha particle then collides with the ^{8}Be nucleus before the nucleus falls apart, a ^{12}C nucleus and another gamma ray can be generated. The ^{12}C nucleus can then collide with a fourth alpha particle to form an oxygen nucleus, ^{16}O. Elhatisari *et al.*[2] report first-principles simulations of collisions between two alpha particles, a first step towards numerical modelling of ^{12}C and ^{16}O formation.

Why did we have to wait until 2015 to obtain a first-principles description of alpha–alpha scattering? The simple answer is that developing a fundamental understanding of nuclei and their interactions is one of the most complicated problems in science. It involves unravelling the properties of an ensemble of nucleons (protons or neutrons), exerting forces on each other, that emerge from the underlying theory of strong interactions, while also accounting for all the quantum-mechanical laws that govern microscopic objects. The complexity of the numerical calculations needed explodes as the number of nucleons increases. Explaining the dynamic interactions of clusters of nucleons is especially hard. Even the scattering of deuterium (a two-nucleon system) from alpha particles has only recently been described from first principles[6].

Elhatisari and co-authors report a clever way to break through this computational ceiling. They start with a supercomputer-friendly formulation — on a four-dimensional space-time lattice — of chiral effective field theory. This theory[7,8] links the interactions of nucleons to quantum chromodynamics (the underlying theory of the strong force) by describing them as the sum of an infinite number of terms, systematically organized in order of importance so that all but the first few terms can be neglected.

The authors use their formulation to follow the evolution of the wavefunction of mutually interacting nucleons in a pair of alpha particles, although this is easier said than done. It requires a trick[9,10] developed in the 1950s and an aptitude for 'gambling': the problem is reformulated as a much simpler system of independent nucleons interacting with a background of auxiliary particles; the most-likely backgrounds are drawn from a probability distribution. This technique, known as auxiliary-field Monte Carlo, is then used

to 'cool' the two alpha particles, which are initially placed at a distance from each other, to their correct physical state — that is, to a low-energy quantum-mechanical solution of two dynamically interacting nucleon clusters. This process is repeated over and over, each time with a new relative position for the initial pair of alpha particles, again drawn from an appropriate probability distribution, and the resulting physical states are then used to compute the effective interaction experienced by the two particles as a whole. *Voilà*! The ferocious eight-body problem is transformed into a docile two-cluster problem.

After 2 million hours of parallel computations, the alpha–alpha scattering properties obtained by the authors — including the first three interaction terms of the sequence provided by chiral effective field theory — show promising agreement with experimentally obtained values. The calculation is admittedly somewhat shy of the accuracy required to make quantitative predictions for nucleosynthesis and stellar evolution. Improvements should be made by including the next term in the sequence, investigating the dependence of the results on the spacing of the space-time lattice, and doing other precision tests. Extensions to enable the treatment of three-cluster dynamics[11] are also required before the method can be applied to the triple-alpha process.

Most impressively, Elhatisari *et al.* have devised a first-principles method for simulating scattering and reactions in which the number of computing operations is proportional to the square of the number of nucleons, and therefore grows relatively slowly. Scattering between alpha particles and ^{12}C, and the conversion of these particles to ^{16}O — a problem that is only four times as difficult as alpha–alpha scattering using the authors' approach — are now within reach of simulations. Furthermore, analogous methods could be used to solve other puzzles. For example, predictive calculations of hyperon–neutron scattering could help to settle whether or not 'strange' particles can exist in the cores of neutron stars[12], thus providing insight into the phases of dense nuclear matter that can exist. But that is another story. ■

Sofia Quaglioni *is in the Nuclear and Chemical Sciences Division, Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551-0808, USA.*
e-mail: quaglioni1@llnl.gov

1. Rutherford, E. *Phil. Mag.* **47,** 109–163 (1899).
2. Elhatisari, S. *et al. Nature* **528,** 111–114 (2015).
3. Salpeter, E. E. *Astrophys. J.* **115,** 326–328 (1952).
4. Hoyle, F. *Astrophys. J. Suppl.* **1,** 121–146 (1954).
5. Buchmann, L. R. & Barnes, C. A. *Nucl. Phys. A* **777,** 254–290 (2006).
6. Hupin, G., Quaglioni, S. & Navrátil, P. *Phys. Rev. Lett.* **114,** 212502 (2015).
7. Weinberg, S. *Phys. Lett. B* **251,** 288–292 (1990).
8. Weinberg, S. *Nucl. Phys. B* **363,** 3–18 (1991).
9. Stratonovich, R. L. *Dokl. Akad. Nauk SSSR* **115,** 1097–1100 (1957).
10. Hubbard, J. *Phys. Rev. Lett.* **3,** 77–78 (1959).
11. Romero-Redondo, C., Quaglioni, S., Navrátil, P. & Hupin, G. *Phys. Rev. Lett.* **113,** 032503 (2014).
12. Lonardoni, D., Lovato, A., Gandolfi, S. & Pederiva, F. *Phys. Rev. Lett.* **114,** 092301 (2015).

ECOLOGY

Ecosystem vulnerability to ocean warming

Analysis of the temperature ranges occupied by marine species finds that the vulnerability of ecological communities to global warming may depend more on organismal physiology than on the magnitude of change. SEE ARTICLE P.88

DEREK P. TITTENSOR

Human communities have already begun to develop and execute plans for adapting to climate change[1]. Ecological communities are equally vulnerable, and human intervention is required to alleviate pressures and minimize the risks of biodiversity loss and species extinction. Much of our attention so far has focused on predicting the responses of individual species. But is it possible to anticipate how entire ecosystems will respond and reconfigure as the land and oceans warm? In this issue, Stuart-Smith *et al.*[2] (page 88) construct a metric of community vulnerability for marine environments that is based on the physiology of individual species as well as on external environmental conditions. Their findings challenge previous assumptions that the magnitude or rate of warming is the best predictor of ecological change.

All species have a thermal niche — the temperature range in which they can survive. But, in reality, most do not occupy all sites within this range because other constraints, such as competition and food availability, limit them further. The range of temperatures over which a species actually lives is its 'realized' thermal niche. Stuart-Smith *et al.* use two large species-occurrence databases to construct realized thermal niches for almost 4,000 reef fish and marine macroinvertebrate species, by comparing observations of the animals' occurrence with data on sea surface temperature at those locations.

Each realized thermal niche has a midpoint, and the mean of these midpoints for all individuals in an ecological community is called the community thermal index (CTI; Fig. 1).

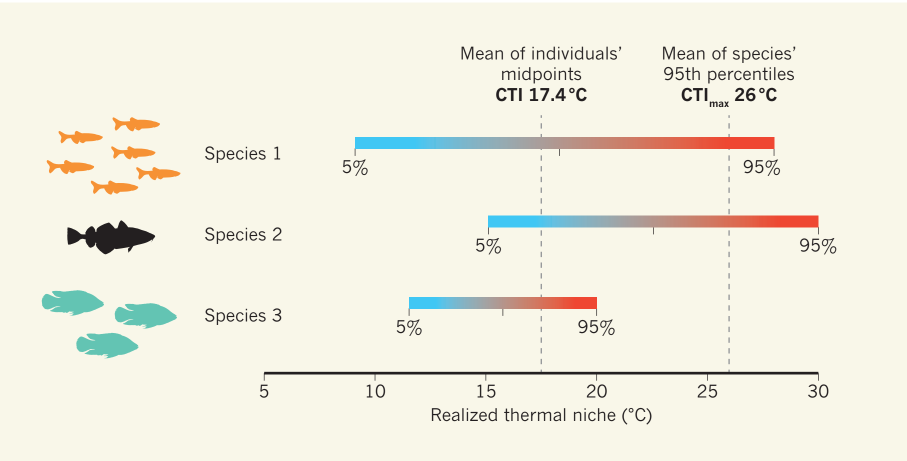


Figure 1 | Community thermal indices. Each species has a realized thermal niche — the range of temperatures over which it can live in a given community. The community thermal index (CTI) is the mean midpoint of these thermal niches for all individuals in a community. Stuart-Smith *et al.*[2] have extended this metric to the CTI_{max}, which is a measure of the mean of the upper end (95th percentile) of the realized thermal limits of the species in the community. The CTI_{max} allows calculation of a vulnerability metric — the proportion of species in a community that has an upper thermal limit lower than a given summer sea surface temperature. In this example, if the future summer sea surface temperature is 25 °C, the vulnerability metric will be 0.33, because one of the three species in the community (species 3) has an upper thermal limit below that temperature.

Although the CTI is not a new concept (see ref. 3, for example), Stuart-Smith *et al.* calculate it across a global range of marine communities. They then compare it with observed temperatures to determine a 'thermal bias' — the discrepancy between the CTI and the mean annual sea surface temperature. This indicates whether a community is weighted towards species adapted to warmer environments (a positive bias) or to cooler ones.

The authors find that most communities are associated with a thermal bias. This is perhaps unsurprising, but Stuart-Smith *et al.* also detect an intriguing large-scale biogeographical pattern, with the thermal bias not randomly scattered around ocean temperature (see Fig. 1 of the paper[2]). Moving from CTIs to examining the thermal-niche midpoints that underpin them, the authors find that most species are associated with either a temperate or tropical midpoint, with noticeably fewer species having midpoints at subtropical temperatures (there is also a third group of invertebrates that have a subpolar midpoint). Such faunal clustering is responsible for the nonlinear distribution in thermal bias observed in these marine communities. This pattern awaits testing for consistency in other taxa, and the potential mechanisms that establish it require further exploration. The assumption that a species is 'optimized' for its thermal midpoint also needs further experimental or empirical verification, although the authors provide evidence to support it.

Of course, a thermal bias in a community may not represent anything more than an assortment of species with wide thermal ranges, indicating low susceptibility to warming. To move beyond an index of bias and towards an estimate of vulnerability, a measure that accounts for the upper limit of each organism's temperature tolerance must instead be determined. Stuart-Smith *et al.* define the upper limit as the 95th percentile of its thermal distribution; thus, for a species with a 95th percentile of 28 °C, 95% of individuals would be found below this temperature. The authors then define a measure for each community, called the CTI_{max}, as the mean of the 95th percentile of each species in the community (Fig. 1).

Sites with a CTI_{max} close to the summer water temperature are likely to contain many species living perilously close to their thermal limits. The authors encapsulate this in a vulnerability metric, which is defined as the proportion of species at each site that have an upper thermal limit lower than the mean summer water temperature. Projecting temperatures forward 100 years to 2115 using climate models from the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)[4], they predict that one-third of surveyed ecoregions will have all species living at temperatures greater than their upper thermal limits — a stark indication that these individuals must move, adapt or perish.

Will all individuals of those species in these ecoregions die? Not necessarily. Just as correlation does not imply causation, vulnerability does not imply extinction. Many species may have greater plasticity or ability to respond to change than we anticipate. Stuart-Smith and colleagues' study also does not take into account the complex interactions between species, perturbations of which may propagate in unpredictable and complex ways. Other potential biasing factors include the authors' use of the IPCC's most-extreme climate scenario (RCP8.5), and the fact that some other anthropogenic impacts will act synergistically. And as communities reorganize, species may move in as well as out and total species richness thus be unaffected — or even increased.

Nonetheless, the CTI_{max} takes into account organismal physiology rather than just levels or rates of environmental warming, and as such may move us a step closer towards understanding the effects of warming on entire assemblages. Indeed, Stuart-Smith *et al*. find that the sites projected to lose the most species are those with a more negative thermal bias, rather than those with a high magnitude of warming. This suggests that using environmentally based metrics of warming (see ref. 5, for example), without taking species characteristics into account, may be insufficient for characterizing vulnerability. An obvious next step is to test Stuart-Smith and colleagues' approach with other taxa to see whether similar patterns emerge. However, under-sampling of species ranges could give the impression of an artificially narrow niche. The picture provided is only as good as the data underlying it, which may limit broader application of this metric.

The response of ecological communities to climate change is undeniably more complex than any single value can reveal. Suites of metrics are used by international policymakers to track the status of biological communities in response to anthropogenic change[6], and metrics that encompass biological traits may add valuable information. In conjunction with modelling efforts that incorporate species interactions (see, for example, refs 7 and 8), a scaffolding of understanding, or at least plausibility, can be constructed. Stuart-Smith *et al.* have contributed a tool that could help us to reach this goal. But in a world in which marine and terrestrial ecosystems face accelerating pressures[9], our ability to respond, protect and sustain remains precarious. ■

Derek P. Tittensor *is at the United Nations Environment Programme World Conservation Monitoring Centre, Cambridge CB3 0DL, UK, and at Dalhousie University, Halifax, Nova Scotia, Canada.*
e-mail: derek.tittensor@unep-wcmc.org

1. Rosenzweig, C. & Solecki, W. *Global Environ. Change* **28,** 395–408 (2014).
2. Stuart-Smith, R. D., Edgar, G. J., Barrett, N. S., Kininmonth, S. J. & Bates, A. E. *Nature* **528,** 88–92 (2015).
3. Devictor, V. *et al. Nature Clim. Change* **2,** 121–124 (2012).
4. IPCC. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Pachauri, R. K. *et al.*) (IPCC, 2015).
5. Loarie, S. R. *et al. Nature* **462,** 1052–1055 (2009).
6. Secretariat of the Convention on Biological Diversity. *Global Biodiversity Outlook 4* (2014).
7. Fernandes, J. A. *et al. Global Change Biol.* **19,** 2596–2607 (2013).
8. Harfoot, M. B. J. *et al. PLoS Biol.* **12,** e1001841 (2014).
9. Tittensor, D. P. *et al. Science* **346,** 241–244 (2014).

This article was published online on 11 November 2015.

METABOLISM

Inflammation keeps old mice healthy

Immune cells called regulatory T cells accumulate in fat during ageing. The anti-inflammatory activity of these cells worsens age-associated defects in metabolism, in contrast to its effect in obesity. **See Letter p.137**

IVAN MAILLARD & ALAN R. SALTIEL

Body fat undergoes extensive and frequent remodelling, as changes in blood-vessel development, connective tissue, the number and size of fat cells, and other features allow fat to store or release the energy that the organism needs. But this adaptive process can become harmful in stressful conditions. Maladaptive remodelling in obese rodents and humans is associated with larger fat cells and chronic inflammation in fat, leading to insulin resistance, type 2 diabetes and cardiovascular complications[1]. Attempts to disentangle the various components that cause this dysregulation of metabolism have revealed a coordinated inflammatory circuit involving interactions between several cell types[1]. On page 137 of this issue, Bapat *et al.*[2] add a new twist to the story, providing evidence that, contrary to what might be expected, age-associated metabolic changes may be

regulated by a different mechanism from those associated with obesity.

Specialized immune cells called regulatory T cells (T_{reg} cells) suppress the inflammatory immune responses driven by white blood cells. The importance of T_{reg} cells for the functioning of the immune system is highlighted by the fact that mammals that lack the transcription factor FOXP3 — which controls the development, maintenance and function of T_{reg} cells — develop multi-organ autoimmune disease. Although many different immune cells are present in fat, fat-resident T_{reg} cells (fT_{reg} cells) have attracted attention as potential modulators of local inflammation[3,4].

The T_{reg} cells first enter fat immediately after birth, accumulate during ageing and acquire a molecular signature characterized by the expression of several factors[5–8], including the transcription factor PPAR-γ, which controls differentiation, and the protein subunit ST2 (a receptor for the protein IL-33), which has been implicated in the development of fT_{reg} cells. Furthermore, some T-cell-receptor proteins (which recognize structures called antigens during an immune response) are preferentially expressed in fT_{reg} cells over other T_{reg} cells, suggesting that certain fT_{reg} populations proliferate in response to antigens[3–5]. Although much remains to be learnt, data from animal models of obesity suggest that fT_{reg} cells protect against inflammation and metabolic dysfunction[3,5–8].

Bapat *et al.* analysed fT_{reg} cells in ageing mice, rather than in models of obesity, and turned the tables on previous assumptions about these cells. Understanding ageing-associated metabolic dysregulation is vital, because age is a major risk factor for insulin resistance and diabetes. The authors observed a striking age-related accumulation of fT_{reg} cells. But unexpectedly, when the authors depleted the fT_{reg} population by deleting PPAR-γ in these cells, they found that the ageing mutant mice gained less weight than their wild-type counterparts — they accumulated less body fat and more lean weight, ate more and burnt more calories than age-matched controls.

Every metabolic parameter analysed by the authors was better in these mice. Fasting glucose and insulin levels decreased, as did insulin resistance. By contrast, pharmacological expansion of the fT_{reg} population increased levels of insulin resistance and other parameters of metabolic dysregulation.

The researchers found that fT_{reg} depletion was associated with increased local levels of the pro-inflammatory signalling molecule TNF-α, consistent with increased fat inflammation. Depletion was also correlated with a reduction in fat-cell size and with reduced expression of collagen genes — evidence of improved metabolic activity and beneficial fat remodelling. Ageing fT_{reg} cells maintained their molecular signature and their ability to suppress immune responses, indicating that their function had not been subverted. Together, these data demonstrate that fT_{reg} accumulation plays a part in age-associated metabolic dysregulation, and that at least some aspects of the inflammatory response suppressed by fT_{reg} cells are favourable in ageing (Fig. 1).

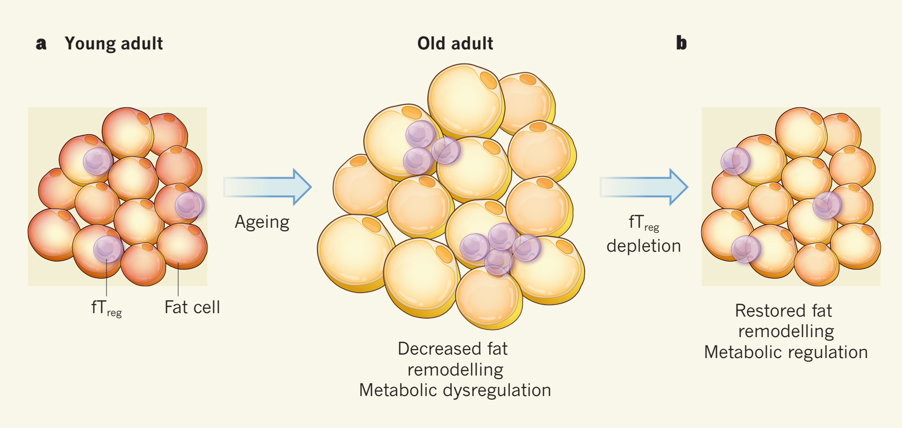


Figure 1 | Regulatory T cells impair metabolic regulation in old fat. a, During ageing, immune cells called fat-resident regulatory T (fT_{reg}) cells accumulate in body fat and decrease local inflammation (red colouring). This age-associated accumulation correlates with increases in metabolic dysregulation and in the size of the fat cells, and with a decrease in the ability of fat to remodel — the process by which fat undergoes morphological alterations in response to changing nutrient demands. **b**, Bapat *et al.*[2] depleted the fT_{reg} population in ageing mice. This increased inflammation, reduced the size of fat cells and improved fat remodelling, thereby improving metabolic regulation.

These surprising results contrast with previous observations, which indicated that fT_{reg} numbers decrease in obese mice, and that fT_{reg} expansion improves metabolic health without affecting weight[3,5–8]. Although Bapat *et al.* focused on ageing, some of their results are notably divergent from previous data. For instance, the authors found no evidence that fT_{reg} cells modulate metabolism in a high-fat-diet model of obesity (although there were few fT_{reg} cells in the fat, and the consequences of expanding the fT_{reg} population were not tested in this model). Directly contradicting previous data[7], they found that stimulating PPAR-γ activity with the anti-diabetic drug rosiglitazone exerted beneficial metabolic effects in obese mice, even those in which PPAR-γ was deleted in fT_{reg} cells, suggesting that other cell types are crucial targets for this drug.

The reasons for these discrepancies remain to be investigated. Perhaps they represent differences in experimental design, or in the populations of commensal bacteria found in mice used at different institutions. It will also be essential to evaluate whether the effects of fT_{reg} cells observed in mice apply to humans. A complicating factor is that ageing- and obesity-associated metabolic dysregulation often coexist in humans.

Nonetheless, information is accumulating about fT_{reg} cells and inflammation. First, other laboratories have also detected age-dependent T_{reg} accumulation in fat[6,9], although the metabolic impact of fT_{reg} cells in ageing has not previously been measured. Second, a refined understanding of the cells' gene-expression profile is emerging, providing hints about their molecular regulation and tools that could be used to manipulate their numbers and function[3,6–8]. Third, compelling evidence[3,5] supports the existence of antigens in fat that drive expansion of fT_{reg} populations; these are probably presented to fT_{reg} cells in association with the antigen-presenting protein complex, major histocompatibility complex class II. It will be crucial to identify these antigens and other factors that contribute to fT_{reg} accumulation during ageing.

Is there crosstalk between fT_{reg} cells and other immune cells, such as innate lymphoid cells (ILCs)? Recent reports[10,11] have shown that ILCs infiltrate fat, and have indicated that IL-33 activates ILC2s to induce white fat to become heat-producing beige fat. It is possible that IL-33 has two opposing roles: inducing energy expenditure through ILC2s, but ensuring energy conservation through expansion of the fT_{reg} population and by inducing the secretion of molecules such as IL-10 that promote anabolic activity (which increases energy storage in fat)[12]. Thus, fT_{reg} depletion could tip the balance in favour of energy expenditure, causing weight loss and improved metabolic health. However, so far there have been no studies on IL-33 or ILC2s in ageing mice.

Finally, blocking inflammatory pathways in ageing fat cells has been reported to impair metabolic regulation[13], in contrast to the prevailing view but consistent with Bapat and colleagues' data. Identifying beneficial elements of the inflammatory response, and investigating their metabolic effects, will be essential, and may point to evolutionarily conserved features of inflammation that lead to tissue adaptation under abnormal conditions. In fat and elsewhere, different facets of inflammation might have many roles, both good and bad. ■

Ivan Maillard *is at the Life Sciences Institute, the Division of Hematology/Oncology and Department of Internal Medicine, and the Department of Cell and Developmental Biology, University of Michigan, Ann Arbor, Michigan 48109, USA.* **Alan R. Saltiel** *is in the Department of Medicine, University of California, San Diego, La Jolla, California 92093, USA.*
e-mails: imaillar@umich.edu; asaltiel@ucsd.edu

1. Osborn, O. & Olefsky, J. M. *Nature Med.* **18,** 363–374 (2012).
2. Bapat, S. P. *et al. Nature* **528,** 137–141 (2015).
3. Feuerer, M. *et al. Nature Med.* **15,** 930–939 (2009).
4. Winer, S. *et al. Nature Med.* **15,** 921–929 (2009).
5. Kolodin, D. *et al. Cell Metab.* **21,** 543–557 (2015).
6. Cipolletta, D., Cohen, P., Spiegelman, B. M., Benoist, C. & Mathis, D. *Proc. Natl Acad. Sci. USA* **112,** 482–487 (2015).
7. Cipolletta, D. *et al. Nature* **486,** 549–553 (2012).
8. Vasanthakumar, A. *et al. Nature Immunol.* **1w6,** 276–285 (2015).
9. Lumeng, C. N. *et al. J. Immunol.* **187,** 6208–6216 (2011).
10. Brestoff, J. R. *et al. Nature* **519,** 242–246 (2015).
11. Lee, M.-W. *et al. Cell* **160,** 74–87 (2015).
12. Lumeng, C. N. & Saltiel, A. R. *J. Clin. Invest.* **121,** 2111–2117 (2011).
13. Wernstedt Asterholm, I. *et al. Cell Metab.* **20,** 103–118 (2014).

This article was published online on 18 November 2015.

In retrospect

A century of phage lessons

One hundred years after the first description of viruses that infect bacterial cells, the contribution of these bacteriophages to fundamental biology, biotechnology and human health continues unabated and deserves celebration.

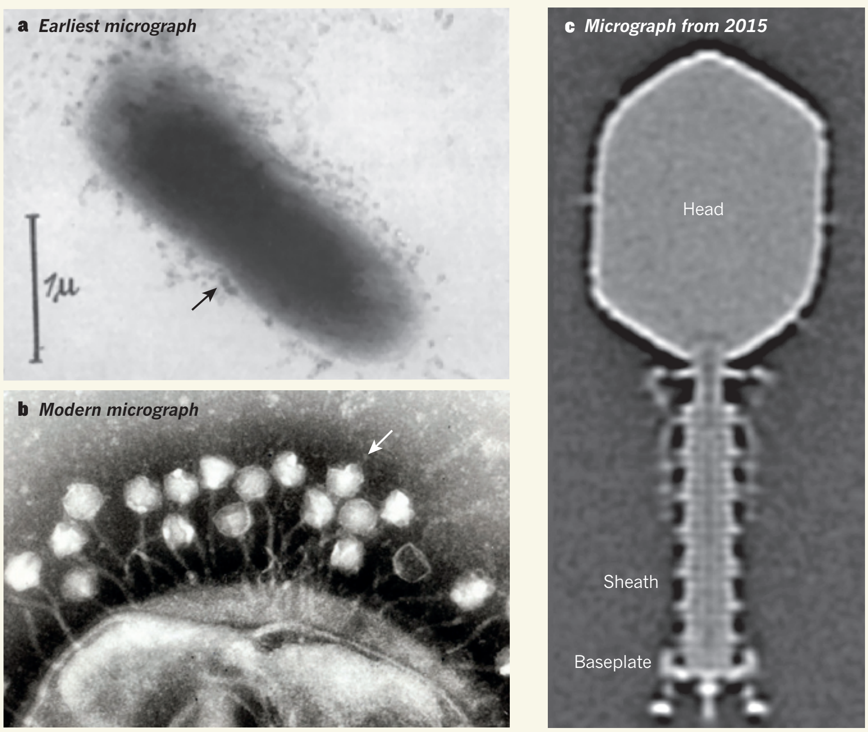
FOREST ROHWER & ANCA M. SEGALL

In 1915, bacteriologist Frederick Twort[1] published the first report of viruses that infect bacteria, replicate there and kill the cells. Since then, studies of these viruses, known as bacteriophages, or more colloquially as phages, have transformed biology. Phages provided the experimental systems and tools for the molecular-biology revolution of the twentieth century, and their rapid growth rates have allowed fundamental principles of ecology and evolution to be tested. We now know that phages are the world's most successful biological entities, being more abundant and genetically diverse than any other life form. Despite their importance, the study of these fascinating entities remains a niche endeavour. Here, we briefly review the history of phage studies, with the hope of inspiring a new generation of phage scientists.

In the early 1900s, most phage scientists were interested in using the viruses as antibacterial agents. This was an era of uncontrolled scientific trials, in which people were injected with phages or the viruses were poured into water wells with the hope of killing pathogenic bacteria, such as those that cause cholera. This line of research dramatically decreased with Alexander Fleming's discovery of antibiotics in 1928. But the concept of 'phage therapy' is currently resurging as antibiotic resistance becomes more of a concern.

Phage science entered the quantitative realm when a network of biologists, biochemists and physicists, known as the Phage Group, used these viruses as models for their pioneering studies of how life works. In 1952, Alfred Hershey and Martha Chase[2] performed a famous experiment in which radiolabelled phages were sheared off bacterial cells using a high-speed blender, helping the researchers to establish that DNA is the genetic material. The discovery of phage-encoded, DNA-manipulating enzymes — such as DNA and RNA polymerases, ligases and endo- and exonucleases — literally catalysed the rise of molecular biology and the biotechnology industry, and phage proteins are now used every day all over the world. Restriction enzymes that protect bacteria from phage infection provided another indispensable tool for molecular biologists. This trend continues today, as can be seen from the revolution in genome editing that has arisen following the discovery of the CRISPR–Cas system, used by bacteria as a defence against phages.

As the genetic code was revealed in the mid-1900s, the sequencing of a complete genome became a major research goal. Phages were attractive targets because of their small genome size and the possibility of making large amounts of DNA for sequencing. Frederick Sanger and colleagues[3] sequenced



A, REF. 15; B, GRAHAM BEARDS/CC-BY-3.0; C, REF. 16

Figure 1 | Bacteriophage in action. Bacterium-infecting viruses were first described[1] in 1915, but it was only in 1940 that the first electron micrographs (**a**) of bacteriophages (arrow) infecting bacteria were published[15]. These early images helped to confirm that the effects attributed to phages were indeed caused by viruses, and not by enzymatic activity. Modern electron microscopy (**b, c**) has produced images of phages that reveal details of phage structure and infection processes (see, for example, ref. 16).

the complete genome of the phage ΦX174 in 1977, decades before any cellular genome was completed. As additional phage genomes accumulated, it became apparent that phages exchange genes and large sections of DNA between individuals[4]. This discovery of horizontal gene transfer changed our understanding of how genetic variability is produced. Marine phage communities were the first to be 'shotgun sequenced', leading to the rise of metagenomics — the sequencing en masse of all members of a community[5].

Understanding phages has contributed to our fundamental understanding of host cells and disease (Fig. 1). When phages integrate into bacterial genomes, they can dramatically change the characteristics of their bacterial hosts — many of the most deadly bacterial pathogens, including *Vibrio cholerae* and *Shigella* and *Salmonella* species, acquire virulence factors through this mechanism. Dissecting the biology of phage replication also uncovered several key host-encoded factors that are needed for the phage life cycle, such as the enzyme DNA gyrase[6] and the 'chaperone' protein complexes GroEL and GroES[7].

When the 'war on cancer' was declared by then US President Richard Nixon in 1971, phage biologists were actively recruited into research on human biology. Building on the knowledge that phages encode some proteins that are similar to those of the host, these scientists looked in the human genome for analogous genes from other viruses. Not only did they find such genes, but they also developed the idea of 'proto-oncogenes' present in our genome that, when mutated, are key drivers of cancer.

Other phage researchers moved into the fields of DNA mutagenesis, repair and recombination, providing the basis for our understanding of cancer today. For example, the understanding[8] that pre-existing mutations can give individual cells growth advantages under different environmental conditions led to the idea that cancer cells harbour dozens of pre-existing mutations that may or may not be related to the actual tumour[9]. With the advent of the AIDS epidemic, phage researchers opened the door to our understanding of how retroviruses integrate into the human genome, and what host proteins are involved[10].

The downside of phage scientists moving into different arenas was a massive decline in phage research from the 1970s onwards. Given that phages are such great anvils for the hammers of biologists, why do many researchers pay them so little attention? One reason might be that, as frequently occurs in any old discipline, the literature is dense and filled with acronyms and a changing nomenclature. To help counter this, we provide some guiding principles on phages.

A first key point is the contribution of phages to biological diversity. There are probably more than 10^{31} phage particles on the planet, with approximately 10 phage particles for every bacterial cell[11]. In humans, the main genetic difference between two individuals is the phages in their gut[12]. Among other roles, these viruses form an adaptable immune system that makes use of hypervariable, immunoglobulin-like protein domains similar to those used by antibodies[13].

The second concept is that phages carry genes encoding proteins that modulate the fundamental physiology of the host, such as metabolism and antibiotic resistance. One fascinating example occurs in photosynthesis by oceanic cyanobacteria[14]. The components of the light-gathering antenna complexes produced by these bacteria are highly labile and decay during phage infection. But the phages can carry genes that encode replacement of the damaged proteins, allowing the bacteria to continue to produce biomass and the phages to produce larger bursts of progeny. Thus, these marine phages contribute to the vast turnover of carbon in the oceans by increasing the efficiency and output of photosynthetic processes.

When phages integrate into bacterial genomes, they can dramatically change the characteristics of their bacterial hosts.

A third phage lesson is that the niche space of any bacterial cell is determined by its phages. The main genomic differences between closely related bacteria derive from integrated phages (prophages) and genomic features, ranging from indels to major rearrangements, that help to guard against phage infection. This never-ending selective pressure exerted on bacteria by their phages is the best-characterized example of the Red Queen hypothesis — that predator and prey species must constantly evolve.

What will the phage future look like? These viruses are relatively easy to synthesize, and their genomes have modular characteristics that appeal to synthetic biologists for engineering biological functions. One hundred years after their discovery, we think that it is time for our fellow biologists to throw off their cell-centric habits and embrace the phage. ■

Forest Rohwer *and* **Anca M. Segall** *are at the Viral Information Institute, Department of Biology, San Diego State University, San Diego, California 92182, USA.*
e-mails: frohwer@gmail.com; asegall@mail.sdsu.edu

1. Twort, F. W. *Lancet* **186,** 1241–1243 (1915).
2. Hershey, A. D. & Chase, M. *J. Gen. Physiol.* **36,** 39–56 (1952).
3. Sanger, F. *et al. J. Mol. Biol.* **125,** 225–246 (1978).
4. Hendrix, R. W., Lawrence, J. G., Hatfull, G. F. & Casjens, S. *Trends Microbiol.* **8,** 504–508 (2000).
5. Breitbart, M. *et al. Proc. Natl Acad. Sci. USA* **99,** 14250–14255 (2002).
6. Nash, H. A. *Annu. Rev. Genet.* **15,** 143–167 (1981).
7. Georgopoulos, C. *Genetics* **174,** 1699–1707 (2006).

NATURE

50 Years Ago

The Architecture of Molecules. By Prof. Linus Pauling and Roger Hayward — "We are now living in an atomic age. In order to understand the world, every person needs to have some knowledge of atoms and molecules." This is the beginning paragraph of a fascinating work of art … The question as to how some understanding of science, however superficial, can be brought to the man and woman in the street has exercised many organizations as well as individuals. At a practical level, of course, it is unnecessary to know anything about electric currents in order to turn a switch and bring on the light. Babies love to do it before they are one year old. But for all too many people science is still magic even when they are twenty-one … Linus Pauling worries. He thinks, quite rightly, that young people ought to want to know why the 'lead' of a pencil comes off on to the paper, what an atom of hydrogen or uranium 'looks like' … Undoubtedly many an arts sixth-former will pick the book off the school library shelf and will learn a great deal by browsing through it.
From *Nature* 4 December 1965

100 Years Ago

The nation's attitude towards science is, I think, largely due to the popular idea that science is a kind of hobby followed by a certain class of people, instead of the materialisation of the desire experienced in various degrees by every thinking person to learn something about innumerable natural phenomena still unsolved; and, having learned, to control and apply them intelligently for the benefit of the human race … It is to the new generation now being educated that we must look for betterment of our position … We must make all education more scientific.
From *Nature* 2 December 1915

8. Luria, S. E. & Delbrück, M. *Genetics* **28,** 491–511 (1943).
9. Vogelstein, B. *et al. Science* **339,** 1546–1558 (2013).
10. Craigie, R. & Bushman, F. D. *Cold Spring Harb. Perspect. Med.* **2,** a006890 (2012).
11. Wommack, K. E. & Colwell, R. R. *Microbiol. Mol. Biol. Rev.* **64,** 69–114 (2000).
12. Reyes, A. *et al. Nature* **466,** 334–338 (2010).
13. Barr, J. J. *et al. Proc. Natl Acad. Sci. USA* **110,** 10771–10776 (2013).
14. Sharon, I. N. *et al. ISME J.* **5,** 1178–1190 (2011).
15. Ruska, H. *Naturwissenschaften* **28,** 45–46 (1940).
16. Hu, B., Margolin, W., Molineux, I. J. & Liu, J. *Proc. Natl Acad. Sci. USA* **112,** E4919–E4928 (2015).

QUANTUM PHYSICS

Getting the measure of entanglement

A property called entanglement entropy helps to describe the quantum states of interacting particles, and it has at last been measured. The findings open the door to a deeper understanding of quantum systems. SEE ARTICLE P.77

STEVEN ROLSTON

A puzzling aspect of quantum mechanics is entanglement: the idea that the combined state of two particles can be completely specified, but that the state of each entangled particle is completely random when measured alone. Entanglement entropy marries the concept of entanglement with that of entropy — the degree of randomness of a system — and has become a useful theoretical tool with which to characterize many-body states in condensed-matter physics. On page 77 of this issue, Islam *et al.*[1] report the first experimental measurement of entanglement entropy in a small system of atoms trapped in a lattice of light, a model of a solid-state system.

Quantum mechanics, the theory of the microscopic world, has many features that run counter to our everyday experiences in a classical world. The possibility of entanglement in a quantum system of two or more particles has been a challenging and stimulating idea for many years. Einstein and his colleagues were famously bothered by the idea that measuring one particle of an entangled pair seemingly instantaneously determined the state of its partner — "spooky action at a distance", as they put it[2]. But the existence of entanglement was made concrete through the theoretical work of the physicist John Bell[3], and experimental tests of Bell's inequalities (constraints derived from Bell's work) have unambiguously verified the quantum-mechanical description of the microscopic world (see ref. 4, for example).

Although an understanding of two-particle entanglement is quite well in hand, there is no specific measure of the amount of entanglement in three or more particles. Yet entanglement has become an important tool for understanding the states of many-body systems. When many particles interact with one another, even through simple interactions, the low-energy quantum states can be surprisingly complicated, with lots of entanglement. Entanglement entropy has become a favoured theoretical measure for categorizing such complex states.

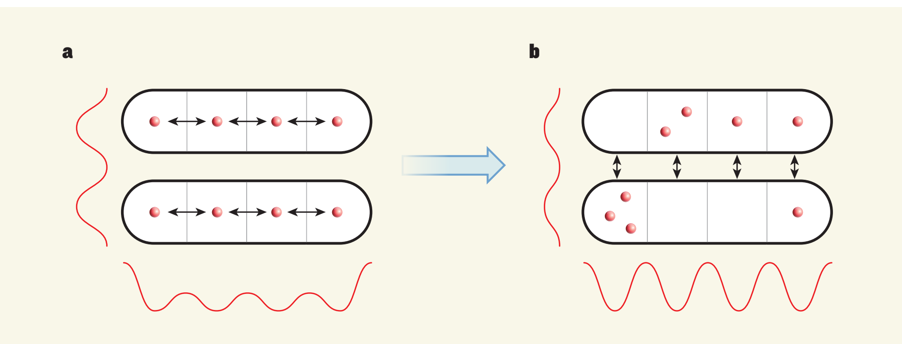


Figure 1 | Probing entropy entanglement in optical lattices. Islam *et al.*[1] report the first experimental measurements of entanglement entropy, a quantity used in theoretical studies to characterize many-body states. **a**, The authors set up two identical systems of four entangled atoms (dots; double-headed arrows indicate horizontal tunnelling), trapped in the potential-energy wells of an optical lattice (an array of interfering laser beams; red lines indicate the optical confining fields). **b**, The confining fields were then adjusted to allow the two four-atom systems to tunnel vertically into one another. The resulting number of atoms in each lattice site contained a signature of each system's state, from which the entanglement entropy can be extracted.

To understand what many-particle entanglement means, let's start by considering a non-entangled system. If I create a system that has N particles, each in an identical state independent of their $N-1$ neighbours, then its many-body description is simple, and measuring one particle or partitioning the sample has little impact on the overall system. Not that such states are uninteresting — this is a good description of a state of matter called a Bose–Einstein condensate, for example. Similarly, if each particle is in its own different state, with no relationship to its neighbours, then measurement or partitioning has no global effect.

But if the particles are entangled with one another, either pairwise or in a more complex fashion, then measurement of one particle affects the state of other particles. Entanglement entropy measures the increase in entropy (which can be thought of as increased randomness) that occurs if we partition such a system[5]. Identifying emergent, complex, lowest-energy states of seemingly simple systems of interacting particles is a particularly challenging task, for which entanglement entropy can be used to understand the nature of the state and to probe its 'quantumness'.

Until now, entanglement entropy has been a purely theoretical construct in condensed-matter physics, because it is difficult to partition a solid-state system and measure its constituents. Islam *et al.* have performed the first such measurements using two identical copies of a small system of four atoms trapped in an optical lattice (an array of interfering laser beams). If the potential-energy 'landscape' of the optical lattice is not too deep, the particles can tunnel from one site to the next and feel the presence of their neighbours. This leads to a many-body state that exhibits entanglement. But if the lattice is deep, the particles act as individuals, and are free of entanglement.

The authors performed their experiment in a quantum gas microscope[6], in which a single layer of an optical lattice is generated just below a high-resolution optical microscope. When Islam *et al.* relaxed some of the optical confining fields, the two copies of the four-atom systems could tunnel into one another and, through quantum interference (the Hong–Ou–Mandel effect[7]), leave a signature of their state in the number of atoms in each lattice site (Fig. 1). The authors simply counted the atoms using the microscope and extracted the entanglement entropy (the second-order Rényi entanglement entropy[5], for those in the know) from the number of atoms. In this way, they show that their four-atom system can have less entropy as a whole than when it is partitioned,

something that is not possible without entanglement, nor in any classical system.

As the first measurement of its kind, this is a milestone. But as with any first, the experimental techniques involved have been pushed to their limits. The 'many-body' systems therefore consist of only four particles, primarily limited by how well the two copies interfere. With improvements, it should be possible to study larger numbers of atoms and more-interesting interacting systems. An intriguing possibility would be to measure higher-order entanglement — the current experiment measures second-order entanglement, whereas *n*th-order entanglement would require *n* interfering copies. This would give further access to the entanglement spectrum, which yields complete knowledge of the quantum state of a system.

Understanding the way in which complex many-body states appear and evolve in systems out of equilibrium is a hot topic in condensed-matter physics, because much of our world is not in equilibrium. This is an especially interesting question in closed systems for which there is no means of driving the system to a thermal equilibrium. Entanglement entropy will be a crucial tool for understanding non-equilibrium systems, and Islam and colleagues' experimental approach is easily adaptable to such studies. The authors' proof-of-principle experiment also opens the door to a greater understanding of the role of entanglement in complex many-body systems through direct experimental observations. Given that both entanglement and entropy are sometimes perplexing concepts, the ability to acquire tangible information about them in the laboratory will certainly benefit their study. ■

Steven Rolston *is at the Joint Quantum Institute, Department of Physics, University of Maryland, College Park, Maryland 20742, USA.*
e-mail: rolston@umd.edu

1. Islam, R. *et al. Nature* **528,** 77–83 (2015).
2. Einstein, A., Podolsky, B. & Rosen, N. *Phys. Rev.* **47,** 777–780 (1935).
3. Bell, B. *Physics* **1,** 195–200 (1964).
4. Hensen, B. *et al. Nature* **526,** 682–686 (2015).
5. Horodecki, R., Horodecki, P., Horodecki, M. & Horodecki, K. *Rev. Mod. Phys.* **81,** 865–942 (2009).
6. Bakr, W. S. *et al. Science* **329,** 547–550 (2010).
7. Hong, C. K., Ou, Z. Y. & Mandel, L. *Phys. Rev. Lett.* **59,** 2044–2046 (1987).

BRAIN CANCER

Tumour cells on neighbourhood watch

The discovery of microtube structures that link tumour cells in some invasive brain tumours reveals how these cancers spread, and how they resist treatment. SEE ARTICLE P.93

HARALD SONTHEIMER

Hardly any diagnosis is as distressing as that of a primary brain tumour. These tumours, often known as gliomas, are a varied group that originate from immature stem cells or from glial cells, which support and protect neuronal networks throughout the brain[1]. Gliomas proliferate uncontrollably, destroying surrounding brain tissue and causing profound neurological damage. They are responsible for around 14,000 deaths each year in the United States alone[2], and are almost always deadly, owing to their resistance to radiation therapy and ability to infiltrate healthy brain tissue. In this issue, Osswald *et al.*[3] (page 93) shed light on what confers these destructive abilities.

Gliomas have long frustrated neurosurgeons[4], because cancerous cells invade the surrounding brain before diagnosis is possible, making surgical removal of these tumours inefficient. The tumours move into the brain through extracellular spaces[5], often following the outside of blood vessels or nerve tracts — in contrast to most other cancers, which disseminate through the blood or lymphatic systems. The invading cells must be killed if treatment is to be successful, and so patients typically undergo aggressive radiation therapy in combination with chemotherapy.

Many gliomas, including the most malignant varieties, resist both radiation and chemotherapy. But a small subgroup called oligodendrogliomas, which harbour deletions in two chromosomal regions dubbed 1p and 19q, respond well to radiation treatment and carry a more favourable prognosis[6]. Osswald and colleagues set out to determine what accounts for this difference in radiation sensitivity.

The authors labelled patient-derived glioma cells taken from a variety of tumours before transplanting them into the brains of mice. Using *in vivo* microscopy, they visualized tumour growth and invasion for up to one year through a window implanted in the animals'

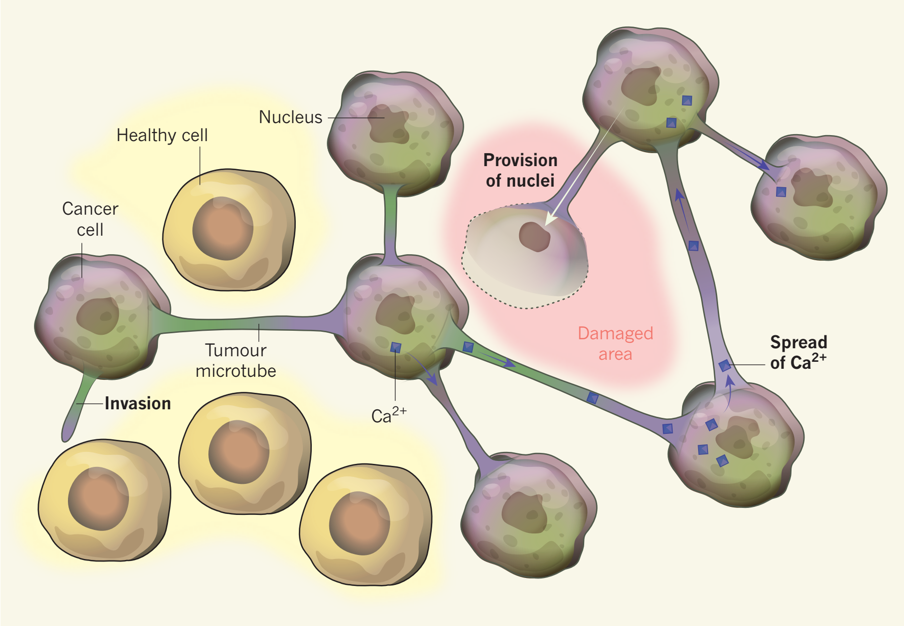


Figure 1 | A network of neighbours. Osswald *et al.*[3] report that, in some types of brain tumour, structures called microtubes connect tumour cells, allowing them to act as a single, organism-like unit. Tumour microtubes facilitate invasion into healthy brain tissue. They permit the spread of toxic molecules such as calcium ions (Ca^{2+}) that build up during radiation therapy, allowing the whole unit to share the burden of toxicity. Furthermore, if tumour tissue is surgically removed, newly synthesized nuclei donated by tumour cells can travel down the microtube to the cell-free site to form new tumour cells.

heads. Invading glioma cells without the 1p and 19q co-deletion extended long, thin, contractile processes into the surrounding brain tissue. The researchers called these structures tumour microtubes.

The microtubes were rich in the proteins actin and myosin, which are known[7] to propel neuronal growth cones (structures at the tips of developing neurons that project out to seek other cells with which to connect). Some microtubes explored, and eventually invaded, the healthy brain. By contrast, others contacted neighbouring tumour cells, forming cytoplasmic bridges between adjacent cells (Fig. 1). This effectively turned the cells into a single organismal unit (a syncytium). The connections between microtubes from each tumour cell were formed by cytoplasm-filled pores called gap junctions, composed of hexameric Cx43 proteins.

What advantage might the tumour gain from growing as a single organismal unit? Osswald *et al.* found that cells of the syncytium were highly resistant to radiation therapy. In unconnected cells, radiation caused an increase in intracellular calcium-ion levels that triggered cell death. But in cells of the syncytium, Ca^{2+} levels remained more stable, presumably because the extra Ca^{2+} was distributed among all cells. Furthermore, when the nucleus of a connected cell was ablated by laser, the neighbouring cell extended a microtube into the affected tissue to deliver a newly synthesized cell nucleus, thus replacing the dead neighbour with a newly nucleated cell — a remarkable feat. The connected cells can thus be thought of as forming a neighbourhood watch group, protecting one another by sharing toxic exposure and even going as far as replacing dead neighbours with part of themselves.

These results indicate that those gliomas that are sensitive to radiation should not be part of a protective neighbourhood. Indeed, when Osswald *et al.* analysed human biopsies from oligodendrogliomas, fewer than 1% contained microtubes and Cx43 expression was almost absent. The authors also found that microtubes and Cx43 were lacking when they transplanted oligodendroglioma-derived cells harbouring the 1p and 19q co-deletion into mice.

Because microtubes are at the heart of the glioma network, the authors next searched for genes that regulate microtube formation. Through Ingenuity Pathway Analysis (a computer-aided method for detecting genes linked to biological traits), they identified *Gap-43* as a candidate. The GAP-43 protein aids neuronal migration and the formation of neural growth cones during development[8]. It was prominently expressed on the invading tips of tumour microtubes, and was conspicuously absent in 1p and 19q co-deleted tumours. Moreover, forced expression of GAP-43 in oligodendroglioma-derived transplants produced highly invasive tumours that acted in protective, microtube-connected networks and resisted radiation. Thus, GAP-43 seems to mediate tumour-microtube formation.

This carefully executed study advances our understanding of brain-tumour growth. For instance, it is now clear that the gap junctions formed by Cx43 are central to the success of glioma neighbourhoods. This protein has been thought to act as a tumour suppressor, preventing cell division in cancers, including gliomas[9]. By contrast, Osswald and colleagues' findings imply that tumour growth is enhanced by the presence of Cx43. This difference can be reconciled when one considers the organism-like growth of the syncytium — in this scenario, the connected cells protect and support one other, allowing the tumour mass to grow even when exposed to radiation.

Although the authors hypothesize that the sharing of Ca^{2+} confers resistance to radiation, many other mechanisms could be at work. Gap junctions allow the passage of many macromolecules between cells, including ATP, amino acids and even microRNAs. One molecule deserving of consideration is the antioxidant glutathione, which readily permeates gap junctions to directly protect cells from radiation damage[7].

From a therapeutic perspective, Cx43 is a challenging pharmacological target. The protein is expressed throughout the body, and is required both for glial transport of cellular metabolites, signals and waste products, and to ensure that heart cells contract in synchrony. Drugs that block Cx43 have been used to heal chronic skin ulcers in humans[10] and to enhance the effectiveness of the chemotherapeutic drug temozolamide in treating gliomas in mice[11]. However, Osswald and colleagues' discovery that GAP-43 is responsible for establishing glioma networks might point to a more effective target for combating these destructive cancers. ■

Harald Sontheimer *is at the Virginia Tech Carilion Research Institute, Glial Biology in Health, Disease & Cancer Center, Roanoke, Virginia 24016, USA.*
e-mail: sontheimer@vt.edu

1. Stiles, C. D. & Rowitch, D. H. *Neuron* **58,** 832–846 (2008).
2. Central Brain Tumour Registry of the United States *Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2004–2008* (CBTRUS, 2012).
3. Osswald, M. *et al. Nature* **528,** 93–98 (2015).
4. Sontheimer, H. in *Diseases of the Nervous System* (ed. Sontheimer, H.) 259–288 (Academic, 2015).
5. Seifert, S. & Sontheimer, H. *J. Physiol. (Lond.)* **592,** 5109–5127 (2014).
6. Chamberlain, M. C. & Born, D. *J. Neurooncol.* **125,** 249–251 (2015).
7. Smith, S. J. *Science* **242,** 708–715 (1988).
8. Denny, J. B. *Curr. Neuropharmacol.* **4,** 293–304 (2006).
9. Naus, C. C. & Laird, D. W. *Nature Rev. Cancer* **10,** 435–441 (2010).
10. Grek, C. L., Rhett, J. M. & Ghatnekar, G. S. *FEBS Lett.* **588,** 1349–1364 (2014).
11. Murphy, S. F. *et al. Cancer Res.* http://dx.doi.org/10.1158/0008-5472.CAN-15-1286 (in the press).

This article was published online on 4 November 2015.

doi:10.1038/nature15743

Managing nitrogen for sustainable development

Xin Zhang[1,2], Eric A. Davidson[3], Denise L. Mauzerall[1,4], Timothy D. Searchinger[1], Patrice Dumas[5,6] & Ye Shen[7]

Improvements in nitrogen use efficiency in crop production are critical for addressing the triple challenges of food security, environmental degradation and climate change. Such improvements are conditional not only on technological innovation, but also on socio-economic factors that are at present poorly understood. Here we examine historical patterns of agricultural nitrogen-use efficiency and find a broad range of national approaches to agricultural development and related pollution. We analyse examples of nitrogen use and propose targets, by geographic region and crop type, to meet the 2050 global food demand projected by the Food and Agriculture Organization while also meeting the Sustainable Development Goals pertaining to agriculture recently adopted by the United Nations General Assembly. Furthermore, we discuss socio-economic policies and technological innovations that may help achieve them.

More than half the world's people are nourished by crops grown with synthetic nitrogen (N) fertilizers, which were made possible in the early twentieth century by the invention of the Haber–Bosch process, which reduces atmospheric nitrogen gas (N_2) to reactive forms of N (ref. 1). A reliable supply of N and other nutrients essential for plant growth has allowed farmers to increase crop production per unit land greatly over the past century, thus promoting economic development, allowing larger populations, and sparing forests that would probably otherwise have been converted to agriculture to meet food demand[2]. Despite this progress, nearly one billion people remain undernourished[3]. In addition, the global population will increase by two to three billion by 2050, implying that demands for N fertilizers and agricultural land are likely to grow substantially[2,4]. Although there are many causes of undernourishment and poverty, careful N management will be needed to nourish a growing population while minimizing adverse environmental and health impacts.

Unfortunately, unintended adverse environmental and human health impacts result from the escape of reactive N from agricultural soils, including groundwater contamination, eutrophication of freshwater and estuarine ecosystems, tropospheric pollution related to emissions of nitrogen oxides and ammonia gas, and accumulation of nitrous oxide, a potent greenhouse gas that depletes stratospheric ozone[5–9] (Fig. 1). Some of these environmental consequences, such as climate change and tropospheric ozone pollution, can also negatively affect crop yields[10,11] and human health[12]. Hence, too little N means lower crop productivity, poor human nutrition and soil degradation[13], but too much N leads to environmental pollution and its concomitant threats to agricultural productivity, food security, ecosystem health, human health and economic prosperity.

Improving nitrogen-use efficiency (NUE)—that is, the fraction of N input harvested as product—is one of the most effective means of increasing crop productivity while decreasing environmental degradation[14,15]. Indeed, NUE has been proposed as an indicator for assessing progress in achieving the Sustainable Development Goals recently accepted by 193 countries of the United Nations General Assembly[16]. Fortunately, we have a large and growing knowledge base and technological capacity for managing N in agriculture[17], and awareness is growing among both agricultural and environmental stakeholder groups that N use is both essential and problematic[15]. This growing awareness, combined with ongoing advances in agricultural technology, is creating a possible turning point at which knowledge-based N management could advance substantially throughout the world. However, improving NUE requires more than technical knowledge. The cultural, social and economic incentives for and impediments to farmer adoption of NUE technologies and best management practices need to be better understood[15].

Here we analyse historical patterns (1961–2011) of agricultural N use in 113 countries to demonstrate a broad range of pathways of socio-economic development and related N pollution. Our analysis suggests that many countries show a pattern similar to an environmental Kuznets curve (EKC), in which N pollution first increases and then decreases with economic growth[18–21]. So far, most EKC analyses have focused on pollution from industrial and transportation sectors[19,22,23]; the present study is one of a few that consider agricultural N pollution in the EKC context[24,25], and apply it globally. However, patterns of N pollution are neither automatic nor inevitable. Socio-economic circumstances and policies vary widely among countries, affecting factors such as fertilizer to crop price ratios and crop mixes, which, as our analysis shows, influence the turning points of the EKC. Although technological and socio-economic opportunities for NUE improvement vary regionally, our analysis shows that average global NUE in crop production needs to improve from ~0.4 to ~0.7 to meet the dual goals of food security and environmental stewardship in 2050.

Patterns of nitrogen pollution

As a useful indicator of potential losses of N to the environment from agricultural soils[26,27], N surplus (N_{sur}; in units of kg N ha^{-1} yr^{-1}) is defined as the sum of N inputs (fertilizer, manure, biologically fixed N, and N deposition) minus N outputs[28,29] (the N removed within the harvested crop products, N_{yield}; Fig. 1). Some of the N_{sur} recycles within the soil, but most N_{sur} is lost to the environment over the long term, because the difference between annual inputs and outputs is usually large relative

[1]Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, New Jersey 08544, USA. [2]Princeton Environmental Institute, Princeton University, Princeton, New Jersey 08544, USA. [3]Appalachian Laboratory, University of Maryland Center for Environmental Science, Frostburg, Maryland 21532, USA. [4]Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey 08544, USA. [5]Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), 75116, Paris, France. [6]Centre International de Recherche sur l'Environnement et le Developpement (CIRED), 94736 Nogent-sur-Marne, France. [7]Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, Georgia 30602, USA.

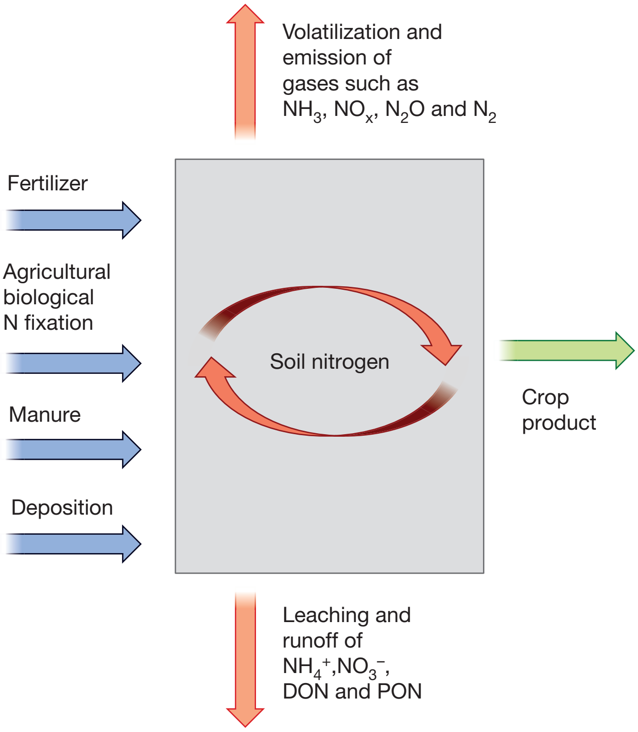


Figure 1 | An illustration of the N budget in crop production and resulting N species released to the environment. Inputs to agriculture are shown as blue arrows and harvest output as a green arrow. NUE is defined as the ratio of outputs (green) to inputs (blue) (i.e. NUE = N_{yield}/N_{input}). The difference between inputs and outputs is defined as N_{sur}, which is shown here as orange arrows for N losses to the environment and as N recycling within the soil (grey box) (that is, $N_{sur} = N_{input} - N_{yield}$). Abbreviations: ammonia (NH_3), nitrogen oxides (NO_x), nitrous oxide (N_2O), dinitrogen gas (N_2), ammonium (NH_4^+), nitrate (NO_3^-), dissolved organic nitrogen (DON) and particulate organic nitrogen (PON).

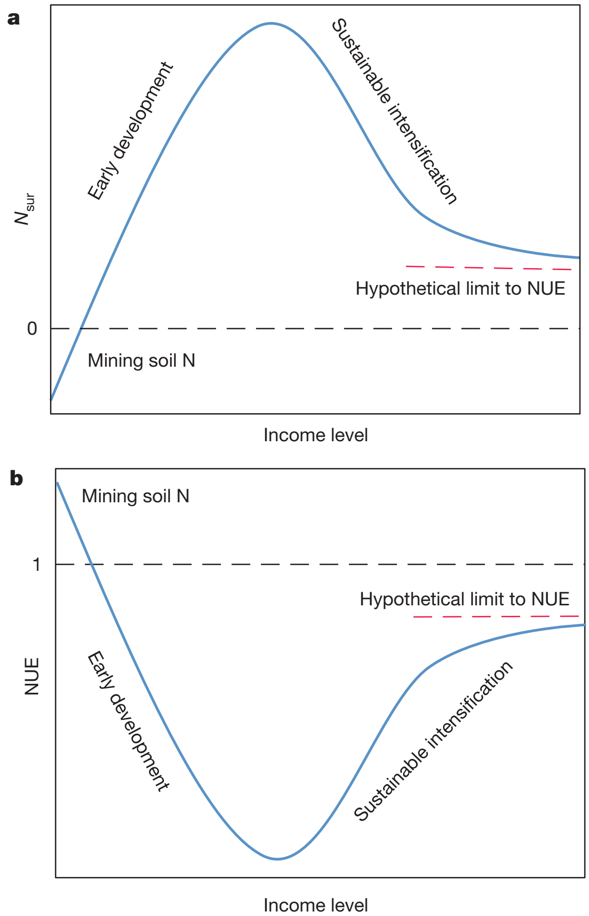


Figure 2 | An idealized EKC for N_{sur} and the related curve for NUE. **a**, The EKC for N_{sur}. **b**, The curve for NUE, which is related to the EKC for N_{sur}. The theoretical limit for NUE (assuming no soil mining of nutrients) is unknown, but no biological system is 100% efficient, so the hypothetical NUE limit is shown as close to but less than unity.

to changes in soil N stocks. The related term of NUE, also called the output–input ratio of N, is mathematically defined as the dimensionless ratio of the sum of all N removed in harvest crop products (outputs or N_{yield}) divided by the sum of all N inputs to a cropland[30,31] (Fig. 1). The N_{sur}, NUE and N_{yield} terms can serve as environmental pollution, agricultural efficiency, and food security targets[32,33], respectively, which are inherently interconnected through their mathematical definitions[33] (that is, $N_{sur} = N_{yield}\left(\frac{1}{\text{NUE}} - 1\right)$, see Supplementary Information section 1 for more information) and their real-world consequences (Fig. 1).

Variable turning points on the EKC

As an indicator of the extent of environmental degradation, N_{sur} aggregated to a national average for all crops is closely related to income growth, mainly in two contrasting pathways as follows. On the one hand, increasing income enables demand for more food consumption[33], which can increase both the land area devoted to agriculture and the intensity of agricultural production and consequently results in more N lost to the environment. On the other hand, increasing income is often accompanied by a societal demand for improved environmental quality, such as clean water and clean air, and is also accompanied by access to advanced technology[18,19]. Consequently, governments may impose regulatory policies or offer subsidies and incentives targeted at reducing local or regional N pollution, and farmers may adopt more efficient technologies.

Therefore, we hypothesize that N_{sur} follows a pattern similar to the EKC: N_{sur} increases with income growth and the quest for food security at early stages of national agricultural development (first phase), but then decreases with further income growth during a more affluent stage (second phase), eventually approaching an asymptote determined by the theoretical limit of the NUE of the crop system (third phase, Fig. 2).

Sustainable intensification of agriculture has been advanced as the key to achieving the second phase of the EKC, including use of cultivars best adapted to the local soil and climate conditions, improved water management, balancing N application with other nutrient amendments, precision timing and placement of fertilizer and manure applications to meet crop demands, the use of enhanced-efficiency fertilizers, and support tools to calculate proper dosing[14,17,34]. While N_{sur} is the EKC environmental degradation indicator, the mathematical relationship between N_{sur} and NUE results in nearly mirror images in Fig. 2 (although see Supplementary Information section 1 for a discussion of situations in which N_{sur} and NUE can both increase simultaneously).

Of the three phases of the N_{sur} trend, it is the second phase of sustainable intensification with increasing affluence that is of greatest contemporary interest. The first phase of agricultural expansion is well documented[30,31], and the third phase cannot yet be evaluated. So far, no country has yet approached the third phase, nor do we know how close to 100% efficiency the use of N inputs could become. For the first phase, as incomes rise, virtually all countries initially increase fertilizer use, N_{yield}, and N_{sur} while NUE decreases[30,31]. To test the existence of the second phase, we examine whether the relationship between gross domestic product (GDP) per capita and N_{sur} breaks away from the linearly (or exponentially) increasing trend and follows more of a bell-shaped pattern over the long term.

We tested the existence of a sustainable intensification phase (or an EKC pattern) with a five-decade record (1961–2011) of N_{sur} and GDP per capita[28,35–40] with a fixed effects model[41–43] across 113 countries for which sufficient data were available and a regression model for each individual country[18,44–46] (see sections 1 and 2 in the

Supplementary Information). The fixed effects model shows a significant quadratic relationship between GDP per capita and N_{sur} ($P < 0.001$, Supplementary Table 9). Regressions between GDP per capita and N_{sur} for each individual country fall into five response types (examples of each group are shown in Fig. 3). Of the 113 countries, 56 countries (group 1) show bell-shaped relationships between N_{sur} and GDP per capita, indicating that N_{sur} increased and then levelled off or decreased as economic development proceeded, as expected for an EKC (two examples are illustrated in Fig. 3a). Those 56 countries account for about 87% of N fertilizer consumption and about 70% of harvested area of all 113 countries. These data provide support for an EKC pattern for N pollution from agriculture, although as we show below, the potential causes of EKC shapes and turning points are complex. Furthermore, for 28 of the 56 countries, by 2011 the rate of increase in N_{sur} had only slowed or levelled off and had not yet actually decreased, indicating likely but still uncertain conformance with an EKC (Supplementary Tables 5 and 6).

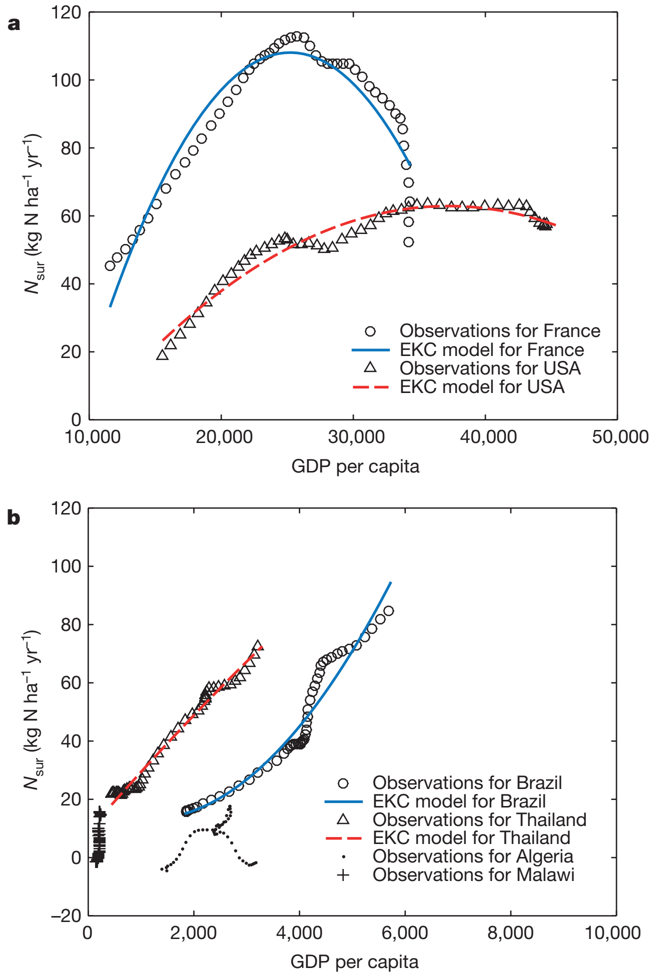


Figure 3 | Examples of historical trends of the relationship between GDP per capita and N_{sur}. The observations are the record of annual N_{sur} smoothed using a ten-year window for each country; the model results are the outcome of the regression using the following model: $Y = a + bX + cX^2$, where the dependent variable Y is the country's N_{sur} and the independent variable X is the country's GDP per capita. We categorized the 113 countries into five groups, based on the significance (that is, P value) and sign of the regression coefficients b and c (see Supplementary Information sections 2.1 and 3.1). **a**, France and USA are examples of group 1, which have significantly negative c ($P_c < 0.05$ and $c < 0$), thus indicating that N_{sur} has started to level off or has declined; **b**, Brazil, Thailand, Malawi and Algeria are examples of groups 2–5, which increase nonlinearly, increase linearly, have no significant correlation ($P_b > 0.05$ and $P_c > 0.05$), or have a negative surplus in 2007–2011, respectively (see Supplementary Tables 5 and 6). The results for all countries can be found in the figures in the Supplementary Information.

Countries with a linear or accelerating increase in N_{sur} (group 3 and most countries in group 2) as GDP per capita grew have not yet approached an EKC turning point (for example, Fig. 3b), but could still follow an EKC in the future as their N input growth slows and NUE increases. Most countries showing an insignificant ($P > 0.05$) relationship between N_{sur} and GDP per capita (group 4) or with a negative N_{sur} (group 5) have had such little income growth and use so little N that the EKC concept cannot be evaluated yet owing to limited change in the country's GDP per capita (for example, Fig. 3b).

Classic empirical studies on EKC, such as Grossman and Krueger (ref. 19), have been criticized because of concerns regarding statistical analyses of time series data that may be non-stationary[47–49]. Therefore, we examined the stationarity of our data (Supplementary Table 7) and used the Autoregressive Distributed Lag modelling approach (ARDL)[50], which is the most frequently used method for the co-integration test in EKC empirical studies published in the last decade[43], to test co-integration on a subset of the data. The ARDL regression models showed the same long-term relationships between N_{sur} and GDP per capita as presented above for all tested countries (Supplementary Table 8). The application of the ARDL method in EKC studies has also been criticized recently for including the quadratic term in the co-integration test, and some new methods have been proposed[51,52]. Further evaluation is needed on the limitations and performance of the ARDL and newly proposed methods for EKC analyses.

Another common criticism of the EKC concept is that the turning point for transitioning to declining environmental degradation is highly variable among pollutants and among countries[18,53,54]. Consistent with those observations, no specific value of GDP per capita was a good predictor of turning points for N_{sur} on the EKC among countries in the present study. For example, N_{sur} in Germany and France started to decline when GDP per capita reached about US$25,000 in the 1980s, while N_{sur} in the USA levelled off and started to decline more recently when GDP per capita reached about US$40,000. Our analysis also shows that countries have widely differing values of NUE and N_{sur} even when yields are similar. Some of this variation is probably due to underlying biophysical conditions, such as rainfall variability and soil quality, which influence crop choices, yield responses, and NUE. However, cultural, social, technological, economic and policy factors also probably affect the turning points on the EKC trajectory of each country.

The turning point in European Union (EU) countries appears to have been reached at least in part owing to policies[55]. Beginning in the late 1980s and through the early 2000s, increases in NUE and decreases in N_{sur} in several EU countries coincided with changes in the EU Common Agricultural Policy, which reduced crop subsidies, and adoption of the EU Nitrates Directive, which limited manure application rates on cropland[56,57]. Relying mostly on volunteer approaches in the USA, the levelling off and modest decrease in N_{sur} since the 1990s is largely the result of increasing crop yields while holding N inputs steady (Fig. 4a), which has resulted from improved crop varieties, increased irrigation and other technological improvements[57,58]. A few state regulatory programmes have required nutrient management plans, placed limitations on fertilizer application dates and amounts, and required soil and plant testing, with varying degrees of success[58–60]. Concerns about water and air quality, estuarine hypoxic zones, stratospheric ozone depletion, and climate change have also stimulated many outreach efforts by governments, fertilizer industry groups, retailers, and environmental organizations to provide farmers with information, training and innovative financial incentives to improve NUE voluntarily[15,59,61,62].

Fertilizer to crop price ratios

Policy can affect NUE not only through regulation and outreach, but also by affecting prices at the farm gate. The ratio of fertilizer to crop prices, R_{fc}, has been widely used in combination with data on yield responses to fertilizer application to advise farmers on fertilizer application rates that yield optimal economic returns[63–65]. In addition to influencing fertilizer application rates, R_{fc} also affects farmer decisions regarding their

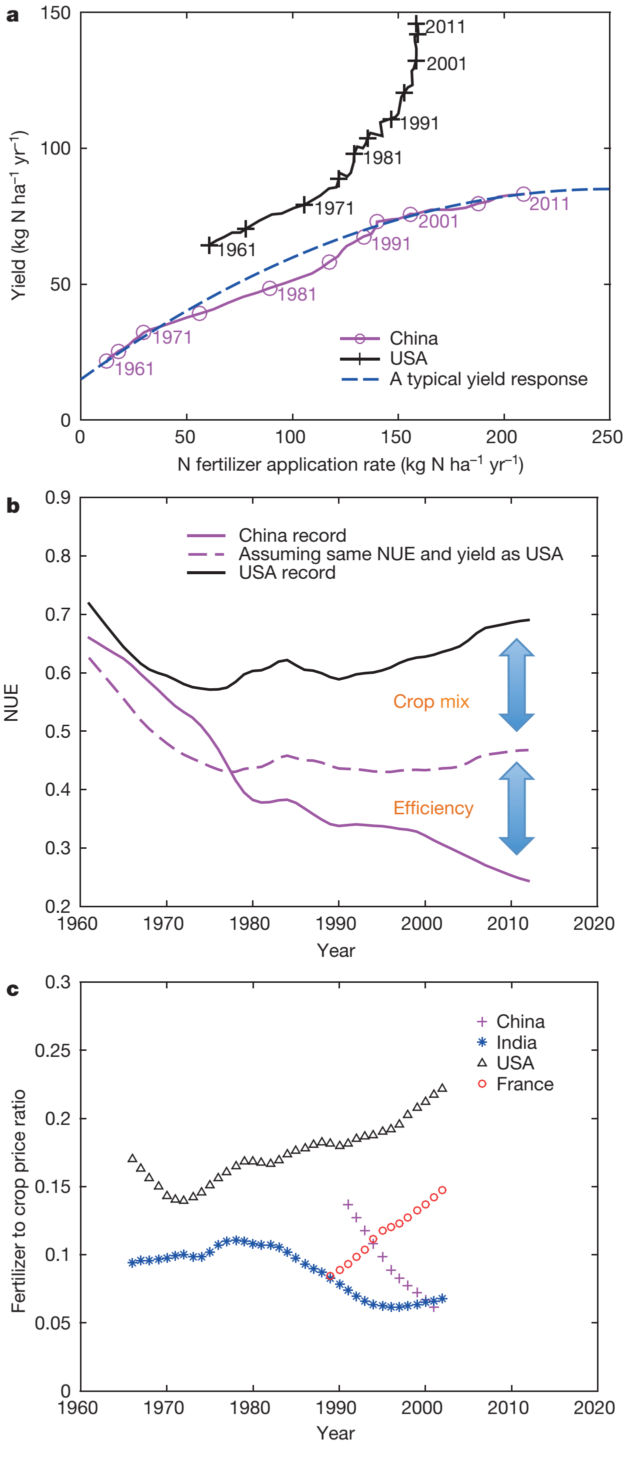


Figure 4 | A comparison of historical trends. a, Nationally averaged annual fertilization rates and yields of maize in China and the USA. **b**, NUE averaged across crops in China and the USA. **c**, Fertilizer to crop price ratios for China, India, USA and France. The dashed blue line in **a** shows a typical yield response function for maize based on fertilizer response trials[33,63], which demonstrates diminishing return in yield as N inputs increase. Note that the historical trend for China follows a pattern similar to a typical yield response function, indicating that further increases in N application rates will result in diminishing yield returns in China. In contrast, maize yield has increased in the USA since 2001 without increasing nationally averaged N input rates, suggesting that the yield improvement has been achieved by adopting more efficient technologies or management practices that shift the yield response curve upwards[33]. The dashed pink line in **b** shows what the NUE in China would be if it achieved NUE values realized in the USA for all crops, but with the crop mix of China. The gap between the dashed pink line and the black line (USA record) is the difference in NUE between countries that is attributable to the differences in crop mixes. The fertilizer to crop price ratio shown in **c** is determined by the N price of urea divided by the N price of maize product (see section 1.6 in Supplementary Information for data sources and methodologies). The data are smoothed using a ten-year window.

choice of technologies and practices for nutrient management, all of which affect NUE and N_{sur} (ref. 33). We tested whether the influence of R_{fc} appears at the national level using two methods: one examines the correlation coefficient of R_{fc} and NUE for individual countries, and the other applies a fixed effects model to all data to test the correlation between R_{fc} and NUE with and without including GDP per capita and crop mix (see section 2.3 in Supplementary Information). Because both the fertilizer and crop prices are 'at the farm gate', they include the effects of government subsidies[35]. The results for maize, for which the most data are available, indicate that the fertilizer to maize price ratio is positively correlated with NUE using both statistical approaches (Supplementary Table 12). We also found that maize prices are linearly correlated with the prices of most major crops, so we infer that the fertilizer to maize price ratio is likely to be a good index for the long-term trend of R_{fc} for all crops. Indeed, we found a statistically significant ($P < 0.001$) positive correlation between historical values of R_{fc} for maize and the NUE aggregated for all other crops. Moreover, this correlation is still statistically significant ($P < 0.001$) after adjusting for the effect of GDP per capita and crop mix (Supplementary Table 11).

Increases in R_{fc} since the 1990s, in both France and the USA (Fig. 4c), coincided with increases in NUE (ref. 57) and may have affected the EKC turning point. At the other extreme, both China and India have had declining values of R_{fc} (Fig. 4c), owing to heavily subsidized fertilizer prices[25,66]. Fertilizer subsidies reached US$18 billion in China in 2010 (ref. 66). Rates of N inputs have now reached levels of diminishing returns for crop yield in China (Fig. 4a), and China has the largest N_{sur} and one of the lowest nationally averaged NUE values in the world (Table 1). The very low R_{fc} in China incentivizes farmers to attempt to increase crop yield by simply adding more N or by choosing more N-demanding cropping systems (for example, change from cereal production to greenhouse vegetable production[67]) instead of adopting more N-efficient technologies and management practices.

Not all fertilizer subsidies are inappropriate. Where infrastructure for producing and transporting fertilizers is poor, as is the case for most of Africa, the cost can be so high that fertilizer use is prohibitively expensive for smallholder farmers, resulting in low yield and small, even negative N_{sur} (soil mining). In these cases, there is room for fertilizer subsidies to increase N inputs, because significant increases in N inputs could be absorbed and greatly increase crop yields without much immediate risk of N pollution[68–70]. When properly designed, temporary fertilizer subsidies structured to build up the private delivery network and with a built-in exit strategy can be an appropriate step[71]. The longer-term question for these countries will be whether they can 'tunnel through' the EKC by shifting crop production directly from a low-yield, high-NUE status to a high-yield, high-NUE status. This shift will require leapfrogging over the historical evolution of agricultural management practices by employing technologies and management practices that promote high NUE before N_{sur} grows to environmentally degrading levels. Acquiring and deploying such technologies, such as improved seed, balanced nutrient amendments, and water management, will require investments in technology transfer and capacity building.

Importance of crop mix

Another factor that may confound EKC trajectories is the mix of crops countries grow over time, which is affected by both demand and trade policies[72]. For example, changing patterns of crop mixes help to explain some of the differences between China and the USA. Since the 1990s an increasing percentage of agricultural land in China has been devoted to fruit and vegetable production, and N application to fruits and vegetables

Table 1 | N budget and NUE in crop production by region and crop in 2010 and projected for 2050

	Current (2010)				Projected (2050)			
	Harvest N (Tg N yr^{-1})	Input N (Tg N yr^{-1})	NUE	Surplus N (Tg N yr^{-1})	Projected harvest N* (Tg N yr^{-1})	Target NUE	Required input N (Tg N yr^{-1})	Resulting surplus N (Tg N yr^{-1})
By region†								
China	13	51	0.25	38	16	0.60	27	11
India	8	25	0.30	18	11	0.60	19	8
USA and Canada	14	21	0.68	7	19	0.75	25	6
Europe	7	14	0.52	7	10	0.75	13	3
Former Soviet Union	4	6	0.56	3	6	0.70	8	2
Brazil	6	11	0.53	5	10	0.70	15	4
Latin America (except Brazil)	7	12	0.52	6	10	0.70	15	4
Middle East and North Africa	3	5	0.48	3	4	0.70	5	2
Sub-Saharan Africa	4	5	0.72	2	9	0.70	13	4
Other OECD countries	1	2	0.52	1	2	0.70	2	1
Other Asian countries	8	19	0.41	11	10	0.60	17	7
Total	74	174	0.42	100	107	0.67	160	52
By crop type‡								
Wheat	13	30	0.42	17	18	0.70	25	8
Rice	11	29	0.39	18	14	0.60	23	9
Maize	13	28	0.46	15	19	0.70	28	8
Other cereal crops	5	9	0.53	4	7	0.70	11	3
Soybean	16	20	0.80	4	24	0.85	28	4
Oil palm	1	1	0.46	1	1	0.70	2	1
Other oil seed	4	10	0.43	6	8	0.70	11	3
Cotton	2	5	0.37	3	3	0.70	5	1
Sugar crops	1	5	0.19	4	2	0.40	4	2
Fruits and vegetables	3	25	0.14	21	5	0.40	11	7
Other crops	5	11	0.41	7	7	0.70	10	3
Total	74	174	0.42	100	107	0.68	157	50

The 2010 record is aggregated from our N budget database (see Supplementary Information section 1 for detailed methodologies and data sources used in developing this database). The 2050 projected harvest N is derived from a FAO projection of crop production to meet a scenario of global food demand[3]. The calculated target NUE values for 2050 are not meant to be prescriptive for particular countries or crops; rather, they are presented to illustrate the types of NUE values that would be needed, given this assumption of food demand[3], while limiting N_{sur} to near the lower bound (50 Tg N yr^{-1}) of allowable N pollution estimated in planetary boundary calculations[78]. Harvest N, input N and surplus N values are rounded to the nearest Tg N yr^{-1}.
*The projected harvest N is based on an FAO scenario[3] for 2050 that assumes a world population of 9.1 billion people and increases in average caloric consumption to 3,200 kcal per capita in Latin America, China, the near East and north Africa, and an increase to 2,700 kcal per capita in sub-Saharan Africa and India. Consumption of animal products increases in developing countries, but differences between regions remain.
†The definitions of the country groups are in Supplementary Table 13.
‡The crop group is defined according to the International Fertilizer Industry Association's report on fertilizer use by crop[38].

now accounts for about 30% of total fertilizer consumption[38,73], with an average NUE of only about 0.10 (which is below the globally averaged NUE for fruits and vegetables of 0.14, and well below the global averages for other major crops; Table 1)[74,75]. At the same time, China has been increasingly relying on imported soybeans, an N-fixing crop that has very low N_{sur} (Table 1)[76]. In contrast, US soybean production has been growing and now accounts for about 30% of the harvested area for crop production (excluding land devoted to production of grasses or crops for feeding livestock) in the USA. While fertilizer subsidies in China probably account for much of the low NUE there, our analysis shows that the difference in crop mix also accounts for nearly half of the NUE difference between China and USA (Fig. 4b).

To address this issue globally, we tested the relationship between NUE and the fraction of harvested area for fruits and vegetables with a fixed effects model for the 113 countries (Supplementary Table 11). The fraction of harvested area for fruit and vegetable production negatively correlates with NUE, and that relationship is still significant ($P < 0.001$) even after adjusting for the effect of GDP per capita.

Meeting the growing challenge

Agriculture is currently facing unprecedented challenges globally. On one hand, crop production needs to increase by about 60%–100% from 2007 to 2050 to meet global food demand[3,77–79]. On the other hand, anthropogenic reactive N input to the biosphere has already exceeded a proposed planetary boundary[5,80], and the increasing demand for food and biofuel is likely to drive up N inputs even further. Therefore, it is critical to establish global and national goals for N use in crop production and to use those goals as reference points to evaluate progress made and guide NUE improvement.

Global and national goals

The planetary boundary for human use of reactive N that can be tolerated without causing unsustainable air and water pollution has been defined in mainly two ways: (1) as the maximum allowable amount of anthropogenic newly fixed N in agriculture that can be introduced into the earth system (62–82 Tg N yr^{-1})[5,80], and (2) as the maximum allowable N_{sur} released from agricultural production to the environment.

Calculations of planetary boundaries according to the first definition require assumptions about nutrient-use efficiency in agriculture. As NUE increases, more N inputs would be manageable while still remaining within air and water pollution limits because more applied N would be taken up by harvested crops. Therefore, rather than focusing on a planetary boundary of allowable newly fixed N, which varies depending on the NUE assumption, we follow the second approach, by estimating what NUE would be needed to produce the food demand

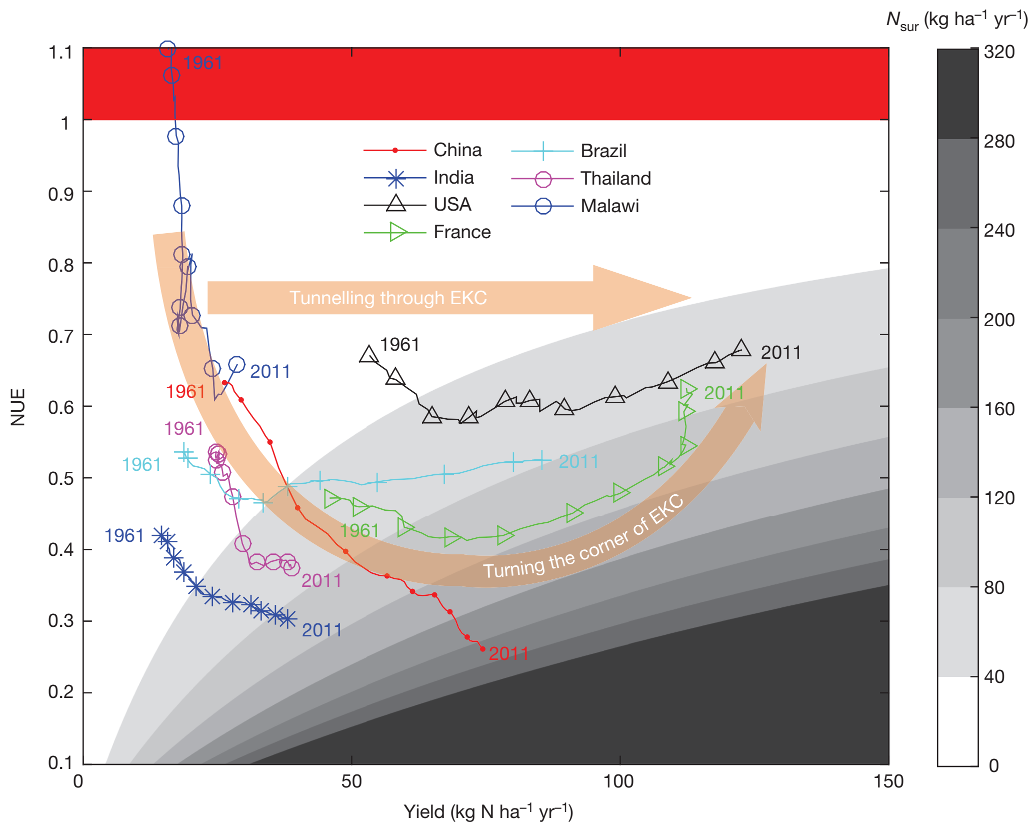


Figure 5 | Historical trends of N_{yield}, NUE and N_{sur}, for a sample of countries examined in this study. The greyscale shows the level of N_{sur}. The area covered in red indicates negative N_{sur}, where the crop production is mining soil N. The data have been smoothed by ten years to limit the impact of year-to-year variation in weather conditions. Curves moving towards the lower right indicate that those countries are achieving yield increases by sacrificing NUE and increasing N_{sur}, whereas curves moving towards the upper right indicate countries achieving yield increases by increasing NUE and resulting in steady or decreasing N_{sur}.

projected for 2050 (ref. 3; Table 1) while keeping N_{sur} within the bounds estimated for acceptable air and water quality. Over 60% of N pollution is estimated to originate from crop production[78], so this is the primary sector that must be addressed to reduce N pollution. From an analysis of the implications of N cycling in several "shared socio-economic pathways"[81], Bodirsky *et al.* (ref. 78) calculated that global agricultural N_{sur} should not exceed about 50–100 Tg N yr^{-1}. Therefore, we use 50 Tg N yr^{-1} as an estimate of the global limit of N_{sur} from crop production.

Meeting the 2050 food demand of 107 Tg N yr^{-1} projected by the Food and Agriculture Organization (FAO, ref. 3) while reducing N_{sur} from the current 100 Tg N yr^{-1} to a global limit of 50 Tg N yr^{-1} (ref. 78) requires very large across-the-board increases in NUE. Globally, NUE would increase from ~0.4 to ~0.7, while the crop yield would increase from 74 Tg N yr^{-1} to 107 Tg N yr^{-1} (Table 1). Recognizing regional differences in crop production and development stage, this average could be achieved if average NUE rose to 0.75 in the EU and USA, to 0.60 in China and the rest of Asia (assuming they continue to have a high proportion of fruits and vegetables in their crop mix), and to 0.70 in other countries, including not dropping below 0.70 in sub-Saharan Africa as it develops (Table 1). Similarly, NUE targets could be established for individual crops, such as improving the global average from 0.14 to 0.40 for fruits and vegetables, and increasing the global average NUE for maize from 0.50 to 0.70 (Table 1).

The challenges in achieving these ambitious goals differ among countries. Figure 5 shows the trajectories of major crop producing countries on the yield–NUE map for the last five decades. The *x* and *y* axes show the two efficiency terms in crop production, NUE and N_{yield}, while the greyscale displays N_{sur}. To compare the nationally averaged field-scale (in units of kg N ha^{-1} yr^{-1}) N_{sur} in Fig. 5 to a global limit of 50–100 Tg N yr^{-1}, the global average N_{sur} target would need to be 39–78 kg N ha^{-1} yr^{-1} across the 2010 harvested area of 1.3 billion hectares. For the examples shown, the USA, France, and Brazil appear to be on this trajectory, although further progress is still needed. In contrast, China and India not only have not yet found an EKC turning point, but also have much ground to make up to reduce their N_{sur} once they turn the corner on their EKC. Although a great challenge, this could also be seen as an opportunity to reduce fertilizer expenditures while increasing agricultural productivity. Malawi, like many sub-Saharan African countries and other least developed countries, has been on a classic downward trajectory of decreasing NUE as it has started to increase N inputs, although evidence from recent years suggests that this decline may have reversed, which would be a necessary first step to tunnelling through the EKC (Fig. 5).

Achieving NUE targets

Achieving ambitious NUE targets while also increasing yields to meet future food demands requires implementation of technologies and management practices at the farm scale, which has been described widely and in considerable detail in the agricultural, environmental, and development literature[17]. Some common principles include the '4Rs' approach of applying the right source, at the right rate, at the right time, in the right place[34]. However, the technologies and management practices needed to achieve the 4Rs vary regionally depending on the local cropping systems, soil types, climate and socio-economic situations. Where improvements in plant breeding, irrigation, and application of available 4R technologies have already made large gains, new technological developments may be needed to achieve further gains, such as more affordable slow-release fertilizers, nitrification and urease inhibitors, fertigation (that is, applying fertilizer via irrigation water), and high-tech approaches to precision agriculture[58].

It is promising that the development and the combination of information technology, remote sensing, and ground measurements will make information about precision farming more readily available, accessible, affordable and site-specific[82]. In many cases, large gains could still be made with more widespread adoption of existing technologies, but a myriad of social and economic factors affecting farmer decision-making regarding nutrient management have only recently begun to receive attention and are critical in improving NUE (ref. 15). Socio-economic impediments, often related to cost and perceived risk, as well as lack of trust in recommendations by agricultural extension agents, often discourage farmers from adopting improved nutrient management practices[59,60,83,84]. Experience has shown that tailoring regulations, incentives, and outreach to local conditions, administered and enforced by local entities, and supported by trust established among local stakeholders improve the success of efforts designed to increase NUE (ref. 15).

Although much of the work must be done at the farm scale, there are important policies that should be implemented on national and multi-national scales. First, improving NUE should be adopted as one of the indicators of the Sustainable Development Goals[16] and should be used in conjunction with crop yield and perhaps other soil health parameters to measure the sustainability of agricultural development. To report reliably on a NUE indicator, countries should be strongly encouraged to collect data routinely on their N management in crop and livestock production. These data should be used to trace trajectories of the three indices of agricultural N pollution, agricultural efficiency and food security targets (that is, N_{sur}, NUE and N_{yield}), as we have done here (Fig. 5) to demonstrate where progress is being made and where stronger local efforts are needed. The data used to construct Fig. 5 have served to demonstrate trends, but both improved data quality and international harmonization of data standards are needed. Regular attention should be given to these trends to establish national and local targets and policies. Just as protocols established by the Intergovernmental Panel on Climate Change permit nations to gauge their progress and commitment for reducing greenhouse gas emissions, protocols for measuring and reporting on a Sustainable Development Goal pertaining to NUE could enable governments to assess their progress in achieving food security goals while maintaining environmental quality.

Second, nutrient management in livestock operations and human dietary choices needs more attention. Here we have focused entirely on crop production, largely because of availability of data, but the N_{sur}, NUE and N_{yield} indices are equally important in livestock management[85]. Indeed, soybeans and some cereals have high NUE as crops, but when fed to livestock, efficient recycling of the N in manure is challenging, resulting in lower integrated NUE for the crop–livestock production system[86]. The crop production scenario used here for 2050 (Table 1) makes assumptions about future dietary choices[3], which are beyond the scope of this study, but we note that future trends in diet will affect the demand for crop and livestock products, the crop mixes grown, and hence the NUE and N_{sur} of future agricultural systems[72].

Third, a similar approach to efficiency analysis would also be valuable for phosphorus (P) fertilizer management, interactions of N and P management, and reducing both N and P loading into aquatic ecosystems[87–90].

Fourth, national and international communities should facilitate technology transfer and promote agricultural innovation. Stronger international collaborations and investments in research, extension and human resources are urgently needed so that knowledge and experience can be shared, creating political and market environments that help to incentivize the development and implementation of more efficient technologies. Technology transfer and capacity building will be needed to enable sub-Saharan African countries to tunnel through the EKC (Fig. 5).

These solutions to improving NUE will require cross-disciplinary and cross-sectorial partnerships, such as: (1) integrating research and development of innovative agricultural technology and management systems with socio-economic research and the outreach needed for such innovations to be socially and economically viable and readily adopted by farmers; (2) analysing the nexus of food, water, nutrients and energy management to avoid pollution swapping (a measure designed to address one pollution problem leads to another; for example, retaining crop residues can reduce nitrogen runoff, but may lead to higher N_2O emission[91]) and to optimize the net benefits to farmers, the environment and society; (3) promoting knowledge and data sharing among private and public sectors to advance science-based nutrient management; and (4) training the next generation of interdisciplinary agronomic and environmental scientists equipped with broad perspectives and skills pertaining to food, water, energy and environment issues.

The EKC has often been described as an optimist's view of a world with declining environmental degradation. Here we have shown that there is evidence—indeed, there is hope—for the EKC pattern of declining N pollution with improving efficiencies in agriculture. However, we have also shown that continuation of the progress made so far is neither inevitable nor is it sufficient to achieve the projected 2050 goals of both food security and environmental stewardship. Turning points and trajectories of national agricultural EKCs will depend largely on agricultural, economic, environmental, educational and trade policies, and these will largely dictate the food and pollution outputs of future agriculture.

Received 5 February; accepted 10 September 2015.
Published online 23 November; corrected online 2 December 2015 (see full-text HTML version for details).

1. Erisman, J. W., Sutton, M. A., Galloway, J., Klimont, Z. & Winiwarter, W. How a century of ammonia synthesis changed the world. *Nature Geosci*. **1,** 636–639 (2008).
2. Foley, J. A. *et al.* Solutions for a cultivated planet. *Nature* **478,** 337–342 (2011).
3. Alexandratos, N. & Bruinsma, J. *World Agriculture towards 2030/2050: the 2012 Revision*. Agricultural Development Economics Division of the Economic and Social Development Department Working Paper No. 12-03, http://www.fao.org/docrep/016/ap106e/ap106e.pdf (Food and Agriculture Organization of the United Nations, 2012).
4. Mueller, N. D. *et al.* Closing yield gaps through nutrient and water management. *Nature* **490,** 254–257 (2012).
5. Steffen, W. *et al.* Planetary boundaries: guiding human development on a changing planet. *Science* **347,** 6223 (2015).
This paper provides the most recent updates on the research under the planetary boundaries framework.
6. Galloway, J. N. *et al.* The nitrogen cascade. *Bioscience* **53,** 341–356 (2003).
This is a classic paper on the many interacting environmental impacts of reactive forms of N as they move through the biosphere.
7. Galloway, J. N. *et al.* Transformation of the nitrogen cycle: recent trends, questions, and potential solutions. *Science* **320,** 889–892 (2008).
8. Reay, D. S. *et al.* Global agriculture and nitrous oxide emissions. *Nature Clim. Change* **2,** 410–416 (2012).
9. Griffis, T. J. *et al.* Reconciling the differences between top-down and bottom-up estimates of nitrous oxide emissions for the U.S. corn belt. *Glob. Biogeochem. Cycles* **27,** 746–754 (2013).
10. Avnery, S., Mauzerall, D. L., Liu, J. & Horowitz, L. W. Global crop yield reductions due to surface ozone exposure: 1. Year 2000 crop production losses and economic damage. *Atmos. Environ.* **45,** 2284–2296 (2011).
11. Robertson, G. P. *et al.* Nitrogen–climate interactions in US agriculture. *Biogeochemistry* **114,** 41–70 (2013).
12. Jerrett, M. *et al.* Long-term ozone exposure and mortality. *N. Engl. J. Med.* **360,** 1085–1095 (2009).
13. Sanchez, P. A. & Swaminathan, M. Hunger in Africa: the link between unhealthy people and unhealthy soils. *Lancet* **365,** 442–444 (2005).
14. Cassman, K. G., Dobermann, A., Walters, D. T. & Yang, H. Meeting cereal demand while protecting natural resources and improving environmental quality. *Annu. Rev. Environ. Resour.* **28,** 315–358 (2003).
15. Davidson, E. A., Suddick, E. C., Rice, C. W. & Prokopy, L. S. More food, low pollution (Mo Fo Lo Po): a grand challenge for the 21st century. *J. Environ. Qual.* **44,** 305–311 (2015).
This paper reports outcomes of an interdisciplinary conference on the technical, social, and economic impediments to improving NUE in crop and animal production systems, and it introduces a series of papers addressing this issue.
16. Leadership Council of the Sustainable Development Solutions Network (SDSN). *Indicators and a Monitoring Framework for Sustainable Development Goals—Revised Working Draft, 16 January 2015*. http://unsdsn.org/resources (SDSN, 2015).
17. Newell Price, J. *et al. An Inventory of Mitigation Methods and Guide to their Effects on Diffuse Water Pollution, Greenhouse Gas Emissions and Ammonia Emissions from Agriculture*. http://www.avondtc.org.uk/Portals/0/Farmscoper/DEFRA%20user%20guide.pdf (Defra Project WQ0106, ADAS and Rothamsted Research North Wyke, 2011).

18. Dinda, S. Environmental Kuznets curve hypothesis: a survey. *Ecol. Econ.* **49,** 431–455 (2004).
19. Grossman, G. M. & Krueger, A. B. Economic growth and the environment. *Q. J. Econ*. **110,** 353–377 (1995).
This was among the first set of studies to provide empirical evidence for the EKC hypothesis.
20. Arrow, K. *et al.* Economic growth, carrying capacity, and the environment. *Ecol. Econ.* **15,** 91–95 (1995).
21. Panayotou, T. *Empirical Tests and Policy Analysis of Environmental Degradation at Different Stages of Economic Development*. Working Paper 238 (Technology and Employment Programme, International Labour Organization, 1993).
22. Cole, M. A., Rayner, A. J. & Bates, J. M. The environmental Kuznets curve: an empirical analysis. *Environ. Dev. Econ.* **2,** 401–416 (1997).
23. Brock, W. A. & Taylor, M. S. in *Handbook of Economic Growth* Vol. 1B (eds Aghion, P. & Durlauf, S.) Ch. 28, 1749–1821 (Elsevier, 2005).
24. Li, F., Dong, S., Li, F. & Yang, L. Is there an inverted U-shaped curve? Empirical analysis of the environmental Kuznets curve in agrochemicals. *Front. Environ. Sci. Eng*. 1–12 (2014).
25. Singh, A. P. & Narayanan, K. Impact of economic growth and population on agrochemical use: evidence from post-liberalization India. *Environ. Dev. Sustain.* **17,** 1509–1525 (2015).
26. van Beek, C., Brouwer, L. & Oenema, O. The use of farmgate balances and soil surface balances as estimator for nitrogen leaching to surface water. *Nutr. Cycl. Agroecosyst.* **67,** 233–244 (2003).
27. Van Groenigen, J., Velthof, G., Oenema, O., Van Groenigen, K. & Van Kessel, C. Towards an agronomic assessment of N_2O emissions: a case study for arable crops. *Eur. J. Soil Sci.* **61,** 903–913 (2010).
28. Bouwman, L. *et al.* Exploring global changes in nitrogen and phosphorus cycles in agriculture induced by livestock production over the 1900–2050 period. *Proc. Natl Acad. Sci. USA* **110,** 20882–20887 (2013).
29. Liu, J. *et al.* A high-resolution assessment on global nitrogen flows in cropland. *Proc. Natl Acad. Sci. USA* **107,** 8035–8040 (2010).
30. Lassaletta, L., Billen, G., Grizzetti, B., Anglade, J. & Garnier, J. 50 year trends in nitrogen use efficiency of world cropping systems: the relationship between yield and nitrogen input to cropland. *Environ. Res. Lett.* **9,** 105011 (2014).
This paper presents the 50-year trend of NUE and the yield response to N input on a country scale.
31. Conant, R. T., Berdanier, A. B. & Grace, P. R. Patterns and trends in nitrogen use and nitrogen recovery efficiency in world agriculture. *Glob. Biogeochem. Cycles* **27,** 558–566 (2013).
This study creates a global N input database by country and several major crops and found no convergence in N use among countries.
32. Brouwer, F. Nitrogen balances at farm level as a tool to monitor effects of agri-environmental policy. *Nutr. Cycl. Agroecosyst.* **52,** 303–308 (1998).
33. Zhang, X., Mauzerall, D. L., Davidson, E. A., Kanter, D. R. & Cai, R. The economic and environmental consequences of implementing nitrogen-efficient technologies and management practices in agriculture. *J. Environ. Qual.* **44,** 312–324 (2015).
This paper develops a bioeconomic model to examine how technological and socioeconomic factors influence farming decisions and the resulting environmental impact.
34. Snyder, C., Davidson, E., Smith, P. & Venterea, R. Agriculture: sustainable crop and animal production to help mitigate nitrous oxide emissions. *Curr. Opin. Environ. Sustain*. **9–10,** 46–54 (2014).
35. Food and Agriculture Organization of the United Nations. *FAOSTAT Online Database* http://faostat.fao.org/ (2015).
36. World Bank Group. *World Development Indicators 2012* http://data.worldbank.org/sites/default/files/wdi-2012-ebook.pdf (World Bank Publications, 2012).
37. Lassaletta, L. *et al.* Food and feed trade as a driver in the global nitrogen cycle: 50-year trends. *Biogeochemistry* **118,** 225–241 (2014).
38. Heffer, P. *Assessment of Fertilizer Use by Crop at the Global Level 2007–2007/08* (International Fertilizer Industry Association, 2009).
39. Monfreda, C., Ramankutty, N. & Foley, J. A. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Glob. Biogeochem. Cycles* **22,** GB1022 (2008).
40. Herridge, D. F., Peoples, M. B. & Boddey, R. M. Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* **311,** 1–18 (2008).
41. Jayanthakumaran, K., Verma, R. & Liu, Y. CO_2 emissions, energy consumption, trade and income: a comparative analysis of China and India. *Energy Policy* **42,** 450–460 (2012).
42. He, J. & Wang, H. Economic structure, development policy and environmental quality: An empirical analysis of environmental Kuznets curves with Chinese municipal data. *Ecol. Econ.* **76,** 49–59 (2012).
43. Al-Mulali, U., Saboori, B. & Ozturk, I. Investigating the environmental Kuznets curve hypothesis in Vietnam. *Energy Policy* **76,** 123–131 (2015).
44. Alam, M. S. & Kabir, N. Economic growth and environmental sustainability: empirical evidence from East and South-East Asia. *Int. J. Econ. Finance* **5,** 86–97 (2013).
45. Diao, X., Zeng, S., Tam, C. M. & Tam, V. W. EKC analysis for studying economic growth and environmental quality: a case study in China. *J. Clean. Prod.* **17,** 541–548 (2009).
46. Song, M.-L., Zhang, W. & Wang, S.-H. Inflection point of environmental Kuznets curve in mainland China. *Energy Policy* **57,** 14–20 (2013).
47. Wagner, M. The carbon Kuznets curve: a cloudy picture emitted by bad econometrics? *Resour. Energy Econ.* **30,** 388–408 (2008).
48. Müller-Fürstenberger, G. & Wagner, M. Exploring the environmental Kuznets hypothesis: theoretical and econometric problems. *Ecol. Econ.* **62,** 648–660 (2007).
49. Chow, G. C. & Li, J. Environmental Kuznets curve: conclusive econometric evidence for CO_2. *Pac. Econ. Rev.* **19,** 1–7 (2014).
50. Pesaran, M. H., Shin, Y. & Smith, R. J. Bounds testing approaches to the analysis of level relationships. *J. Appl. Econ.* **16,** 289–326 (2001).
51. Wagner, M. The environmental Kuznets curve, cointegration and nonlinearity. *J. Appl. Econ*. **30,** 948–967 (2015).
52. Wagner, M. & Hong, S. H. Cointegrating polynomial regressions: fully modified OLS estimation and inference. *Econom. Theory*, http://dx.doi.org/10.1017/S0266466615000213 (2015).
53. Stern, D. I. The rise and fall of the environmental Kuznets curve. *World Dev.* **32,** 1419–1439 (2004).
54. Cavlovic, T. A., Baker, K. H., Berrens, R. P. & Gawande, K. A meta-analysis of environmental Kuznets curve studies. *Agric. Res. Econ. Rev.* **29,** 32–42 (2000).
55. Sutton, M. A. *et al.* (eds) *The European Nitrogen Assessment: Sources, Effects and Policy Perspectives* (Cambridge Univ. Press, 2011).
56. van Grinsven, H. *et al.* Management, regulation and environmental impacts of nitrogen fertilization in northwestern Europe under the Nitrates Directive: a benchmark study. *Biogeosciences* **9,** 5143–5160 (2012).
57. van Grinsven, H. J. *et al.* Losses of ammonia and nitrate from agriculture and their effect on nitrogen recovery in the European Union and the United States between 1900 and 2050. *J. Environ. Qual.* **44,** 356–367 (2015).
58. Ferguson, R. B. Groundwater quality and nitrogen use efficiency in Nebraska's Central Platte River valley. *J. Environ. Qual.* **44,** 449–459 (2015).
59. Osmond, D. L., Hoag, D. L., Luloff, A. E., Meals, D. W. & Neas, K. Farmers' use of nutrient management: lessons from watershed case studies. *J. Environ. Qual.* **44,** 382–390 (2015).
60. Perez, M. R. Regulating farmer nutrient management: a three-state case study on the Delmarva Peninsula. *J. Environ. Qual.* **44,** 402–414 (2015).
61. International Fertilizer Industry Association (IFA). *The Global '4R' Nutrient Stewardship Framework. Developing Fertilizer Best Management Practices for Delivering Economic, Social, and Environmental Benefits*. AgCom/09/44, https://www.ipni.net/ipniweb/portal/4r.nsf/article/global-4r-framework (IFA Task Force on Fertilizer Best Management Practices, IFA, 2009).
62. Davidson, E., Galloway, J., Millar, N. & Leach, A. N-related greenhouse gases in North America: innovations for a sustainable future. *Curr. Opin. Environ. Sust.* **9–10,** 1–8 (2014).
63. Sawyer, J. E. *et al. Concepts and Rationale for Regional Nitrogen Rate Guidelines for Corn*. http://www.extension.iastate.edu/publications/pm2015.pdf (Iowa State University Extension, 2006).
64. Robertson, G. P. & Vitousek, P. M. Nitrogen in agriculture: balancing the cost of an essential resource. *Annu. Rev. Environ. Resour.* **34,** 97–125 (2009).
65. Setiyono, T. D. *et al.* Maize-N: a decision tool for nitrogen management in maize. *Agron. J.* **103,** 1276–1283 (2011).
66. Li, Y. *et al.* An analysis of China's fertilizer policies: impacts on the industry, food security, and the environment. *J. Environ. Qual.* **42,** 972–981 (2013).
67. Ju, X., Kou, C., Christie, P., Dou, Z. & Zhang, F. Changes in the soil environment from excessive application of fertilizers and manures to two contrasting intensive cropping systems on the North China Plain. *Environ. Pollut.* **145,** 497–506 (2007).
68. Hickman, J. E., Tully, K. L., Groffman, P. M., Diru, W. & Palm, C. A. A potential tipping point in tropical agriculture: avoiding rapid increases in nitrous oxide fluxes from agricultural intensification in Kenya. *J. Geophys. Res. Biogeosci*. **120,** 938–951 (2015).
69. Hickman, J. E., Havlikova, M., Kroeze, C. & Palm, C. A. Current and future nitrous oxide emissions from African agriculture. *Curr. Opin. Environ. Sust.* **3,** 370–378 (2011).
70. Zhou, M. *et al.* Regional nitrogen budget of the Lake Victoria Basin, East Africa: syntheses, uncertainties and perspectives. *Environ. Res. Lett.* **9,** 105009 (2014).
71. Jayne, T. S. & Rashid, S. Input subsidy programs in sub-Saharan Africa: a synthesis of recent evidence. *Agric. Econ.* **44,** 547–562 (2013).
72. Billen, G., Lassaletta, L. & Garnier, J. A vast range of opportunities for feeding the world in 2050: trade-off between diet, N contamination and international trade. *Environ. Res. Lett.* **10,** 025001 (2015).
73. Heffer, P. *Assessment of Fertilizer Use by Crop at the Global Level 2010–2010/11*. http://www.fertilizer.org/En/Statistics/Agriculture_Committee_Databases.aspx (International Fertilizer Industry Association, 2013).
74. Shi, W.-M., Yao, J. & Yan, F. Vegetable cultivation under greenhouse conditions leads to rapid accumulation of nutrients, acidification and salinity of soils and groundwater contamination in South-Eastern China. *Nutr. Cycl. Agroecosyst.* **83,** 73–84 (2009).
75. Ju, X.-T. *et al.* Reducing environmental risk by improving N management in intensive Chinese agricultural systems. *Proc. Natl Acad. Sci. USA* **106,** 3041–3046 (2009).
76. Drinkwater, L. E., Wagoner, P. & Sarrantonio, M. Legume-based cropping systems have reduced carbon and nitrogen losses. *Nature* **396,** 262–265 (1998).
77. Searchinger, T. *et al. Creating a Sustainable Food Future: a Menu of Solutions to Sustainably Feed more than 9 billion people by 2050*. World Resources Report 2013-14, Interim Findings (World Resources Institute, 2013).

78. Bodirsky, B. L. *et al.* Reactive nitrogen requirements to feed the world in 2050 and potential to mitigate nitrogen pollution. *Nature Commun*. **5,** 3858 (2014).
79. Tilman, D., Balzer, C., Hill, J. & Befort, B. L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl Acad. Sci. USA* **108,** 20260–20264 (2011).
80. de Vries, W., Kros, J., Kroeze, C. & Seitzinger, S. P. Assessing planetary and regional nitrogen boundaries related to food security and adverse environmental impacts. *Curr. Opin. Environ. Sustain*. **5,** 392–402 (2013).
81. Nakicenovic, N. & Swart, R. (eds) *IPCC Special Report on Emissions Scenarios* (Cambridge Univ. Press, 2000).
82. Mulla, D. J. Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. *Biosystems Eng.* **114,** 358–371 (2013).
83. David, M. B. *et al.* Navigating the socio-bio-geo-chemistry and engineering of nitrogen management in two Illinois tile-drained watersheds. *J. Environ. Qual.* **44,** 368–381 (2014).
84. Weber, C. & McCann, L. Adoption of nitrogen-efficient technologies by US corn farmers. *J. Environ. Qual.* **44,** 391–410 (2014).
85. Powell, J., Gourley, C., Rotz, C. & Weaver, D. Nitrogen use efficiency: a potential performance indicator and policy tool for dairy farms. *Environ. Sci. Policy* **13,** 217–228 (2010).
86. Powell, J. & Rotz, C. Measures of nitrogen use efficiency and nitrogen loss from dairy production systems. *J. Environ. Qual.* **44,** 336–344 (2015).
87. MacDonald, G. K., Bennett, E. M., Potter, P. A. & Ramankutty, N. Agronomic phosphorus imbalances across the world's croplands. *Proc. Natl Acad. Sci. USA* **108,** 3086–3091 (2011).
88. MacDonald, G. K., Bennett, E. M. & Taranu, Z. E. The influence of time, soil characteristics, and land-use history on soil phosphorus legacies: a global meta-analysis. *Glob. Change Biol.* **18,** 1904–1917 (2012).
89. Cordell, D., Drangert, J.-O. & White, S. The story of phosphorus: global food security and food for thought. *Glob. Environ. Change* **19,** 292–305 (2009).
90. Schoumans, O. *et al.* Mitigation options to reduce phosphorus losses from the agricultural sector and improve surface water quality: a review. *Sci. Total Environ.* **468–469,** 1255–1266 (2014).
91. Stevens, C. J. & Quinton, J. N. Diffuse pollution swapping in arable agricultural systems. *Crit. Rev. Environ. Sci. Technol.* **39,** 478–520 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. M. Grossman, Mark W. Watson, G. Chow, Z. Shi, O. Torres-Reyna and Y. Wang for their advice on economic data analysis. We thank E. Shevliakova, F. Gonzalez Taboada and D. R. Kanter for comments. This study was supported by the programme in Science, Technology, and Environmental Policy at the Woodrow Wilson School at Princeton University, the United States Department of Agriculture (grant 2011-67003-30373), and the National Oceanic and Atmospheric Administration, United States Department of Commerce (award NA14OAR4320106). The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration, the US Department of Commerce, or the US Department of Agriculture. This is Scientific Contribution number 5080 of the University of Maryland Center for Environmental Science Appalachian Laboratory.

Author Contributions X.Z., E.A.D., D.L.M. and T.D.S. designed the research. X.Z., T.D.S., and P.D. compiled the N database. X.Z., Y.S. and E.A.D. carried out the statistical analysis. X.Z. and E.A.D. led the writing of the paper with substantial input from D.L.M., T.D.S., P.D. and Y.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.Z. (xzhang@al.umces.edu) or E.A.D. (edavidson@umces.edu).

PERSPECTIVE

doi:10.1038/nature16069

The contentious nature of soil organic matter

Johannes Lehmann[1,2]* & Markus Kleber[3,4]*

The exchange of nutrients, energy and carbon between soil organic matter, the soil environment, aquatic systems and the atmosphere is important for agricultural productivity, water quality and climate. Long-standing theory suggests that soil organic matter is composed of inherently stable and chemically unique compounds. Here we argue that the available evidence does not support the formation of large-molecular-size and persistent 'humic substances' in soils. Instead, soil organic matter is a continuum of progressively decomposing organic compounds. We discuss implications of this view of the nature of soil organic matter for aquatic health, soil carbon–climate interactions and land management.

Soil organic matter contains more organic carbon than global vegetation and the atmosphere combined (Fig. 1). For this reason, the release and conversion into carbon dioxide or methane of even a small proportion of carbon contained in soil organic matter can cause quantitatively relevant variations in the atmospheric concentrations of these greenhouse gases[1]. Moreover, organic matter retains nutrients as well as pollutants in the soil, which improves plant growth and protects water quality[2]. Soils are also an important source of aquatic carbon, with implications for biogeochemical processes in rivers, lakes and estuaries[3]. Despite its recognized importance, there is a widely divergent view of the nature of soil organic matter.

Biological, physical and chemical transformation processes convert dead plant material into organic products that are able to form intimate associations with soil minerals, making it difficult to study the nature of soil organic matter. Early research based on an extraction method assumed that a 'humification' process creates recalcitrant (resistant to decomposition) and large 'humic substances' to make up the majority of soil 'humus' (see Box 1). However, these 'humic substances' have not been observed by modern analytic techniques. This lack of evidence means that 'humification' is increasingly questioned, yet the underlying theory persists in the contemporary literature, including current textbooks[4–6].

Here we argue in favour of a soil continuum model (SCM) that focuses on the ability of decomposer organisms to access soil organic matter and on the protection of organic matter from decomposition provided by soil minerals. Viewing soil organic matter as a continuum spanning the full range from intact plant material to highly oxidized carbon in carboxylic acids[7] represents robust science and will facilitate the way we communicate between disciplines and with the public. Only such an evidence-based approach can allow for the development of mechanistic solutions to climate, water quality and soil productivity issues (Fig. 1). The resulting knowledge should be integrated into conceptual and mechanistic models for the purpose of predicting carbon dioxide emissions from soils in a warming world, as well as of keeping water supplies clean, and of improving and sustaining the ecosystem services of the world's soils. Research aimed at reliable predictions of soil organic matter turnover should focus on investigating its spatial arrangement within the mineral matrix, the fine-scale redox environment, microbial ecology and interaction with mineral surfaces under moisture and temperature conditions observed in soils.

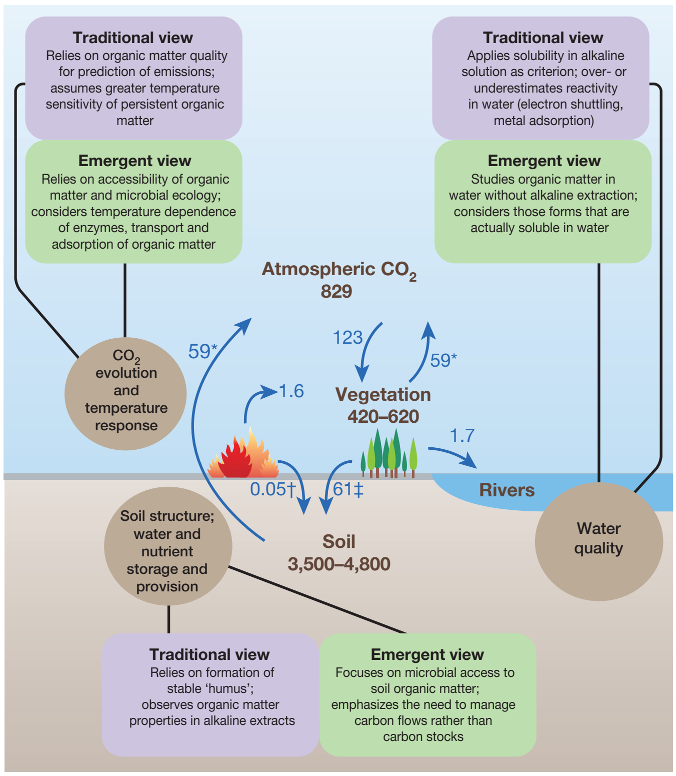


Figure 1 | Traditional and emergent views of the nature of soil organic matter affect how we predict and manage soil, air and water. Traditional 'humification' concepts limit observations of soil organic matter to its solubility in alkaline extracts, unlike the emergent view of organic matter based on solubility in water and its accessibility to microorganisms. Soils are an important source of organic matter in aquatic ecosystems and are responsible for half of the atmospheric carbon recycling. Carbon stocks and flux values are from ref. 1, except where noted otherwise: brown numbers are stocks in Pg C and blue numbers are flows in Pg C yr^{-1}. *Disaggregated value from 119 Pg C yr^{-1} total emissions. †3% of total carbon consumed by fire[104]. ‡Estimate to balance soil carbon exports.

[1]Soil and Crop Sciences, School of Integrated Plant Sciences, College of Agriculture and Life Sciences, Cornell University, Ithaca, New York, USA. [2]Atkinson Center for a Sustainable Future, Cornell University, Ithaca, New York, USA. [3]Department of Crop and Soil Science, Oregon State University, Corvallis, Oregon, USA. [4]Institut für Bodenlandschaftsforschung, Leibniz Zentrum für Agrarlandschaftsforschung (ZALF), Müncheberg, Germany.
*These authors contributed equally to this work.

Historical reliance on an operational proxy

Soil organic matter research is difficult because organic compounds are thoroughly mixed with and often adhere to soil minerals. In arable soil, organic matter typically makes up less than 5% and could historically be discerned only by its dark coloration. Before advanced spectroscopic methods became available in the early 1990s, research on soil organic matter required that the organic phase be separated from the mineral phase through an extraction procedure. The most efficient of these separation procedures in terms of mass extracted[8] is an extraction with alkali (Box 1), which dates back to a report published in 1786 (ref. 9). Although the extraction is incomplete, selective and prone to creating artefacts (Box 1), the procedure became widely adopted and its products universally accepted as experimental proxies for soil organic matter.

Concerns that alkaline preparations are not appropriate representatives of soil organic matter were raised as early as 1888 (ref. 10) and 50 years later it was proposed[11] that 'humic' nomenclature should be dropped because the term relates only to a material obtained by a specific procedure. Unfortunately, these concerns were dismissed rather than disproved. Among the thousands of publications on 'humic substances', not one independently confirms—for example, by direct spectroscopic observation—that the 'humic substances' extracted by alkali are components of organic matter that exist separately in soil environments.

Among the strongest arguments in favour of discarding the notion of 'humic substances' is the absence of any agreement within the broader scientific community on how such materials are defined. 'Humic substances' may be described in the soil sciences in three different ways: strictly operationally according to what can be extracted with an alkaline solution, with further subcategories of 'humic' and 'fulvic' acids as well as unextractable 'humins'; as an existing substance that is not merely an operational construct; or as a combination of the two (Box 1). Different research communities use the same vocabulary with very different connotations, to the point of being contradictory: in soil science, 'humic substances' are thought to have large molecular masses[12]; in the environmental sciences, they are characterized as small fragments[13]; and a classic textbook of aquatic geochemistry describes them as compounds of variable mass and composition[14]. These views have evolved over time, so that now it is not obvious what the term 'humic substances' is intended to convey unless it is explicitly defined. Despite this uncertainty and new insight from modern spectroscopic techniques (Box 2), the products of alkaline extraction continue to be treated as physically existing entities[5,6,15], with research efforts focused on aligning theory with the behaviour and properties of a soil component proxy that is defined solely by solubility at an alkaline pH.

Reconciling models of soil organic matter

At present, three competing models for the fate of organic inputs to soil can be distinguished: (1) classic 'humification', (2) 'selective preservation' and (3) 'progressive decomposition' (Fig. 2).

All three models assume that fragments of plants and soil fauna are first broken up into small pieces at the onset of decomposition. Evidence that such breakdown of dead leaves or roots takes place comes from the observation that the majority of organic matter inputs to soil decays within the first year[16]. It is further known that plant residues must be degraded by enzymes to a relatively small size (typically less than 600 Da) before they can be actively transported across the cell walls of microorganisms[17,18]. In terrestrial ecosystems, so-called exo-enzymes perform this function outside the microorganism[19,20]. Thus, at any time within a living soil, a continuum exists of many different organic compounds at various stages of decay[21], moving down a thermodynamic gradient from large and energy-rich compounds to smaller energy-poor compounds[20].

(1) The 'humification' model is the oldest of the three concepts[22]. In its original definition 'humification' assumes a further transformation or synthesis of the initial decomposition products into large, dark-coloured compounds[12] (Fig. 3). The resulting macromolecules were thought to be rich in carbon and nitrogen structures specific to 'humification', resistant to decomposition[12] and consequently, older than the rest of the soil organic matter. Given the lack of a universally accepted definition of 'humic substances' across disciplines and the lack of evidence for their physical existence independent of the alkaline extraction procedure, it is no surprise that there is no agreement on the processes and pathways of 'humic substance' formation either (Box 2). These 'humic substances' are variously considered to be ecologically useful (providing cation exchange capacity), chemically reactive (interacting with iron, aluminium and other

BOX 1

Traditional approach to the study of soil organic matter

Since first used over 200 years ago, the alkaline extraction technique has undergone many iterations but the principle has remained identical. In its modern version[105], the procedure involves the addition of a sodium hydroxide solution with a very high pH of 13 to a soil sample. At this pH, most oxygen-containing functional groups in organic matter are ionized, making organic compounds bearing such groups much more soluble in water[67]. After adding protons to the solubilized organic materials, a dark solid precipitates that is commonly called 'humic acid'. The organic matter that remains soluble after reacidification is called 'fulvic acid'. The considerable proportion of organic matter that does not respond to the treatment, either for a lack of ionizable functional groups or because it was shielded from the harsh alkaline treatment by mineral protection, is named 'humin'. This multi-step procedure created the need to distinguish several categories of what constitutes soil organic matter. These categories vary widely between authors. The conceptual problem with defining 'humic substances' by an extraction procedure is threefold:

(1) The extraction is always incomplete, leaving 50%–70% of the organic carbon unextracted, which is then defined as the insoluble 'humin' fraction[106]. This precludes the use of the extractable 'humic and fulvic acids' as true representatives of total soil organic matter. The alkaline solution will also extract portions of soil fractions that are not meant to be included in 'humic substances', such as living biomass, simple and identifiable biomolecules (often included as 'non-humic' substances in 'humus'), dissolved organic matter or undecomposed leaves and roots (isolated as particulates). How these separately assessed fractions should be distinguished from the unextracted 'humins' (that are part of 'humic substances') is often unclear. The sum of 'humic' and 'non-humic' substances is defined as 'humus', a term that is sometimes considered to be synonymous to soil organic matter[5,6], sometimes not[19], and is sometimes not used at all[4,5,8,22].

(2) The harsh alkaline treatment at pH 13 ionizes compounds that would never dissociate within the wider soil pH range (pH 3.5 to pH 8.5), giving the resulting 'humic' and 'fulvic' fractions the character of highly selective preparations with an exaggerated chemical reactivity rather than that of true isolates.

(3) The development of this extraction method preceded theory, tempting scientists to develop explanations for the synthesis of materials resembling operationally extracted 'humic substances', rather than to develop an understanding of the nature of all organic matter in soil. Over time, this attempt to mechanistically explain the formation of operationally defined 'humic substances' also led to their definition as synthesis products without the link to the alkaline extraction[5,6].

BOX 2

Critique of the 'humification' model

A consolidated assessment of published evidence (Fig. 3) reveals that secondary synthesis of 'humic substances' facilitated by minerals or enzymes has not been shown to be relevant in natural systems. On these grounds we find it inadvisable to support the classic 'humification' model. Evidence based on isotopic labelling[107] or on the testing of numerous decomposer organisms[108] leaves little doubt that the supposedly recalcitrant 'humic substances' can be decomposed at surprisingly fast rates. The dark colour of 'humic' extracts generated in laboratory experiments[109,110] can be satisfactorily explained by a combination of two processes: the degradation of natural pigments and the accumulation of molecules containing random conjugated bonds (which appear dark in the mixture). Large molecular masses of hundreds to millions of daltons (mostly 10,000–100,000 daltons) reported in early studies[12] have more recently been found to consist of self-assembled aggregates of small compounds mimicking large molecules[13,46,111]. Contrary to many earlier interpretations, the old radiocarbon age of some alkaline extracts[112] is not a valid criterion for the persistence of decomposed organic matter, but merely an indication of when the carbon was fixed by photosynthesis[113]. The chemical structures of so-called polyaromatic carbon compounds (carbon in ring structures) often observed in the extracts are routinely produced by both plants and microorganisms and include melanins, tannins and antibiotics (polyketides)[114,115]. However, these compounds have a clear physiological purpose and are therefore not the products of a random decomposition process. Ubiquitous thermally altered carbon from vegetation fires found in most soils[116–118] is also polyaromatic, and a portion of such compounds is typically extracted in alkaline solution[83,119]. Heterocyclic nitrogen (nitrogen embedded in a carbon ring structure) has been proposed to result from secondary synthesis, but evidence is only available to demonstrate its origin from fires[120] or from artefacts during analyses[15,121]. The glass transition sometimes observed in materials from alkaline extracts[122] has been attributed to 'humification'[123], because glass transition behaviour requires a degree of molecular order. But the glass transition can also be found in many microbial products[124] and fire-altered organic matter[125] (in which the processes are well established).

metals), and—particularly relevant for biogeochemical models—also inherently 'stable' against further decomposition[12]. The suite of hypothetical transformation processes became collectively known as 'humification' and is also called the 'synthesis concept of the genesis of humic substances' or 'secondary synthesis'[12,15,23,24] (Fig. 2).

(2) 'Selective preservation', which is also called preferential decomposition[25], is a newer concept informed by decomposition studies of leaves[26,27] and visible plant fragments in soils[28]. This concept assumes that organic inputs are composed of both labile and relatively recalcitrant compounds[29], the latter being used by microorganisms only when the former are exhausted. However, there is now robust evidence that, under suitable conditions, appropriately adapted decomposer organisms have the ability to decompose even presumably persistent materials more quickly than previously anticipated, including polycondensed aromatics[30], alkanes in soil[31], fire-derived carbon[32], crude oil in sea water[33], and even polyethylene[34]. Also, contrary to previous assumptions[35], the decomposition of presumably recalcitrant lignin is fastest at the early stages of decomposition, as long as it is easily accessible and small organic molecules are available as a source of energy to help mineralize the lignin[36].

(3) In the progressive decomposition model (also called 'biopolymer degradation'[37]; or 'the degradative concept'[15,23]), soil organic matter consists of a range of organic fragments and microbial products of all sizes at various stages of decomposition[7,38] (Fig. 2). Several independent lines of evidence revealed alkali-extracted 'humic substances' to be a mixture of identifiable compounds such as fragments of plants or microorganisms[39–41] that are distributed in different locations of micro-aggregates[42–45], showing no similarity to the 'humic' extract[42], and having small size[46,47]. Upon cell death, materials that are synthesized in the course of microbial anabolism are released into the soil, where they are subject to further degradation. Throughout this process, these materials remain on an energetic downhill trajectory[48], as opposed to the hypothetical 'humic substances' (Fig. 2), whose 'secondary synthesis' would require energy investments for which no thermodynamic rationale has been provided so far[15].

Using recognized chemical, physical and biological controls on soil carbon turnover, the available evidence can reconcile those existing theories into a SCM (Fig. 2). In the SCM concept, organic matter exists as a continuum of organic fragments that are continuously processed by the decomposer community towards smaller molecular size[7,20,21]. The breakdown of large molecules leads to a decrease in the size of primary plant material with concurrent increases in polar and ionizable groups, and thus to increased solubility in water. At the same time, the opportunity for protection against further decomposition increases through greater reactivity towards mineral surfaces and incorporation into aggregates (Fig. 2). Modern analytical tools for the characterization of biomolecules in microbial cells and soils now suggest a direct and rapid contribution of microbial cell walls to soil organic matter protected by interaction with minerals[49,50]. Adsorption may be followed by desorption, exchange reactions with competing organic compounds, and biotic or abiotic degradation. An obvious consequence of microbial involvement in the decomposition process is the direct deposition of microbial cells, cell debris, exopolysaccharides, and root exudates on mineral surfaces.

Only the SCM explains the variations in turnover time of organic compounds through variations in the presence or absence of decomposer organisms and enzymes and the energy they require, through the properties and abundance of mineral surfaces that may protect organic matter, and through the availability of numerous other resources (such as oxygen and nutrients)[51,52]. The vast portfolio of options for variations in carbon turnover dynamics in the SCM provides a full explanation of organic matter properties as observed by contemporary, *in situ* spectro-microscopic techniques[39–42] without invoking 'humification' processes or 'humic substances'. Consequently, the SCM does not require microbial or abiotic generation of recalcitrance through the formation of specific organic compounds and is in agreement with the stated need to focus on spatial arrangement of soil organic matter[53] and environmental control such as temperature, moisture or soil mineralogy[52]. Decomposition pathways, sequences and rates therefore evolve as a specific function of a given soil system. The SCM offers a way forward in modelling soil carbon dynamics and developing soil management that is based on observable evidence, as discussed below.

Environmental relevance

The SCM view of the nature of soil organic matter—which excludes any secondary synthesis of 'humic substances'—has implications for a range of disciplines that build on the science of organic matter properties and changes in soil (Fig. 1). This is all the more important as the 'humic substances' concept is very widely adopted outside the soil sciences, with the majority of publications focusing on 'humic substances' published in journals that do not explicitly cover soil science.

Soil carbon modelling

Soils contain more organic carbon than the atmosphere and vegetation combined[1] and predictions of soil organic matter dynamics could therefore greatly influence forecasts of global climate change. Major soil carbon models such as Century[54] or RothC[55] are built on the premise that soil

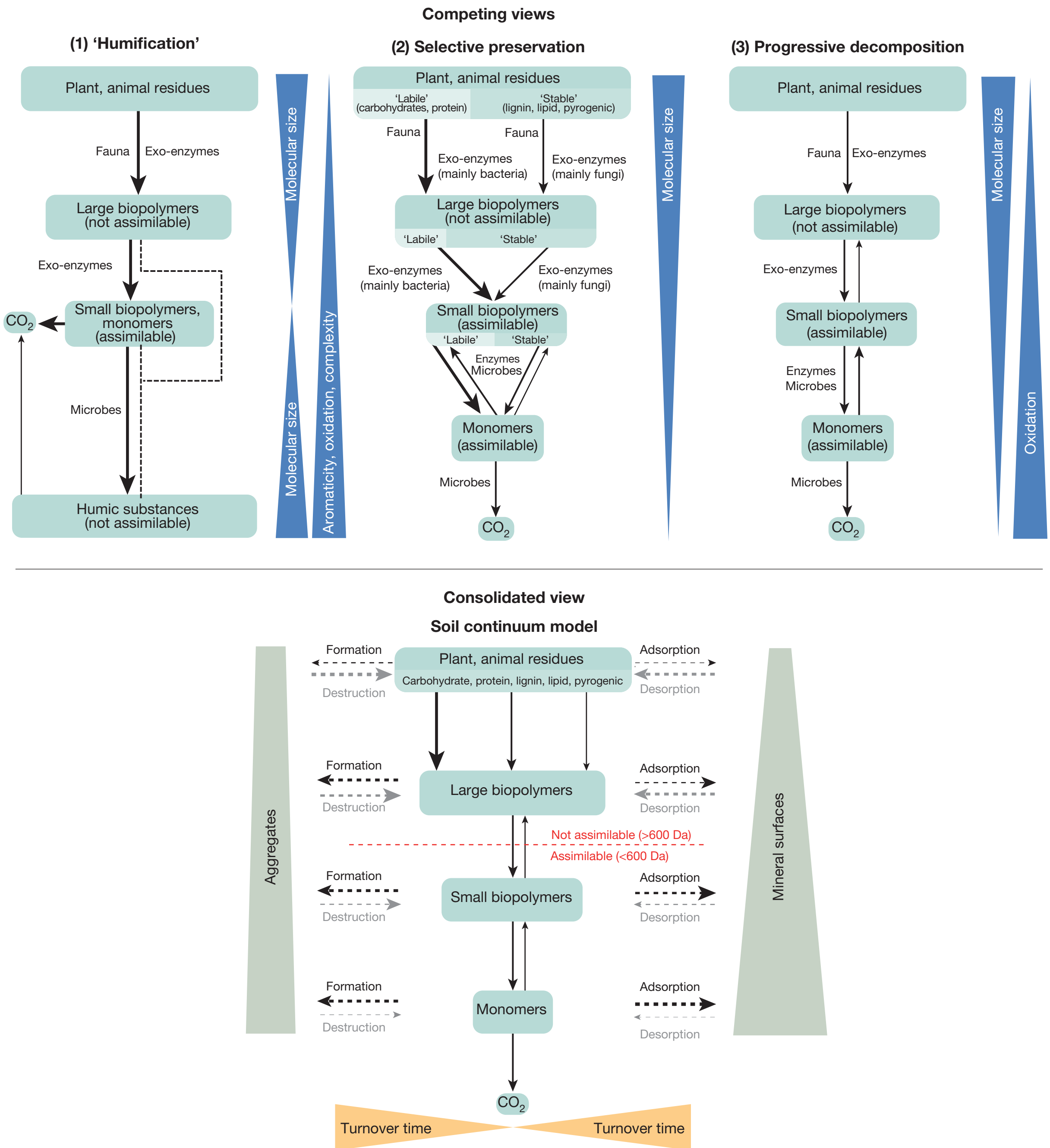


Figure 2 | Reconciliation of current conceptual models for the fate of organic debris into a consolidated view of a SCM of organic matter cycles and ecosystem controls in soil. Classic 'humification' relies on the synthesis of large molecules from decomposition products. Selective preservation assumes that some organic materials are preferentially mineralized, leaving intrinsically 'stable' decomposition products behind. Progressive decomposition reflects the concept of microbial processing of large plant biopolymers to smaller molecules. In the proposed SCM, a continuum of organic fragments is continuously processed by the decomposer community from large plant and animal residues towards smaller molecular size. At the same time, greater oxidation of the organic materials increases solubility in water as well as the opportunity for protection against further decomposition through greater reactivity towards mineral surfaces and incorporation into aggregates. Dashed arrow lines denote mainly abiotic transfer, solid lines denote mainly biotic transfer; thicker lines indicate more rapid rates; larger boxes and ends of wedges illustrate greater pool sizes; all differences are illustrative. All arrows represent processes that are a function of temperature, moisture and the biota present.

organic matter can be divided into pools that have different turnover times. None of these models explicitly represents the characteristic processes of carbon transformation detailed in the SCM, such as adsorption and protection, desorption, and microbial activity. Although carbon movement between pools and their decomposition rates are modified by temperature, texture and moisture, the default turnover rates associated with individual carbon pools are justified by the combined influence of physical protection and an inferred resistance to decomposition that is dependent on substrate quality ('quality' is here used in the sense of molecular composition of the organic matter). Particularly for the 'slow' and 'passive' pools, this inherent resistance to decomposition (recalcitrance) has been understood to be the result of 'humification', with the RothC model explicitly including 'humus' fractions[55]. Lack of mechanistic representation of the decomposition process produces disagreement among models[56] and between model predictions and observational data[57,58].

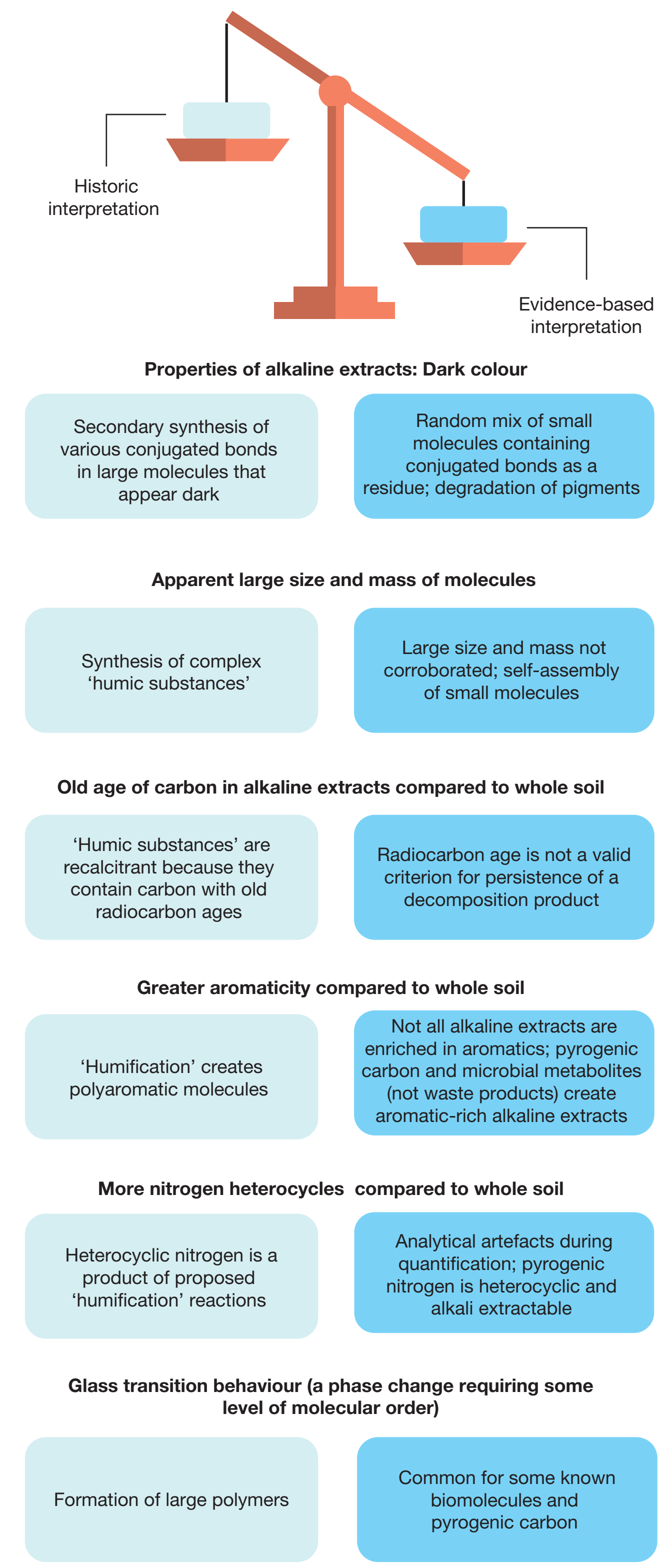


Figure 3 | Weighing up the empirical information supporting either the historic or evidence-based interpretation of the nature of soil organic matter. A consolidated assessment of scientific evidence published over the past two decades provides explanations for the properties of alkaline extracts that do not require invoking the secondary synthesis of 'humic substances'.

The shortcomings become apparent when these models are applied to predict the global warming feedback of soil organic carbon mineralization. Rising temperatures increase microbial activity and a warming atmosphere may therefore lead to greater mineralization of soil organic carbon[59]. The resultant carbon dioxide emissions would then accelerate the greenhouse effect and thereby increase global temperature. Soil organic matter pools with slower turnover are thought to respond more sensitively to climate warming than those with fast turnover[59–61]. The underlying, so-called carbon–quality–temperature theory (CQT theory[62]) combines classical 'humification' theory, that is, the assumption that decomposition creates complex, recalcitrant compounds, with the Arrhenius theory that chemical reactions are faster at higher temperatures[63]. According to CQT theory, the decomposition of a complex substrate requires more enzymatic reactions and a higher total activation energy than a reaction metabolizing a simple carbon substrate, and as a result, would be more sensitive to rising temperatures than the decomposition of a simple carbon substrate. The CQT theory loses much of its explanatory potential for the carbon pools with slow turnover if the decomposition of organic matter is not creating complex and recalcitrant compounds.

Different organic compounds entering the soil have highly varying composition[64] and in isolation (for example, fresh litter) have different turnover and hence temperature responses as a function of their composition[60]. However, this variation is so heavily influenced by environmental and biotic factors after they enter the soil ecosystem that the concept of relying on quality-dependent temperature responses is, in our opinion, obsolete. We propose that future research should concentrate to a much greater extent on the causes of any observed substrate preferences, such as the absence of a decomposer with a matching catabolic toolbox or the lack of a critical resource for the decomposer.

To equip models with more appropriate temperature responses, new approaches need to recognize first the continuum of organic compounds (rather than discreet pools with different turnover times), and second the protection of organic compounds (rather than substrate quality). It is not obvious that merely distinguishing between the mineralization of plant litter on the one hand and degradation products interacting with the mineral matrix[65] on the other will generate better predictive capabilities, simply because they form a continuum. In addition, the full suite of controls on mineralization must be considered, notably temperature–moisture interactions[66]. Mechanistic understanding in this field will be greatly improved if 'humification'-derived assumptions about the molecular structure of the slower-cycling soil carbon pools are replaced by considerations of the processes that render organic decomposition fragments mobile in soil solution. The relevance of binding mechanisms of organic substances to different mineral surfaces is still uncertain[67] and the stability of minerals themselves may change as a result of exposure to organic compounds, such as those released by roots[68].

The laudable efforts to include microbial activity[69] and diversity[70] into soil carbon models to improve climate predictions continue to focus on the quality of organic matter. The development of models built on microbial ecology should omit any emphasis on substrate quality and especially the proposed large 'humified' organic compounds. Observations in soils depleted of plant litter input showed microbial communities adapted to metabolizing simple, small compounds rather than the large and polymeric organic compounds expected for old and persistent soil organic carbon[71]. To predict the responses of soil organic carbon to climate warming, models must move beyond conceptual pools having different turnover times and instead combine soil physical principles into soil biological processes. As recently demonstrated[72], aspects of this combination are already possible when models include the extent to which the mobility of organic fragments in soil water affects accessibility of decomposition products by functionally different groups of microorganisms.

It will next be critical to develop models that provide deeper insight into microbial access to soil organic carbon by including the spatial architecture of the soil[53]. Such model development benefits from spatial data, which are becoming available using imaging analyses in two[42,49] or three dimensions[73]. In a fully developed model, this will require extensive computing capabilities and may only be possible if this research is prioritized or at a time when further computational advances make complex spatial calculations easily accessible and inexpensive. Combining these

approaches within the SCM would provide opportunities to test whether the distance of microorganisms from the organic matter plays as important a part as does the attachment of organic matter to protective mineral surfaces, which constitutes the next frontier in better understanding and prediction of soil organic carbon dynamics.

Aquatic systems

Because soil organic matter is a major source of organic carbon in rivers, lakes and estuaries[3], its persistence and retention is of great interest for closing global carbon budgets[1]. Large proportions of organic carbon in rivers are mineralized and emitted as carbon dioxide[74] or retained in fluvial[75] and oceanic sediments[76]. To date, 'humic substances' as extracted by alkali constitute the organic workhorse that is investigated by the community of aquatic chemists. Continuing this practice of investigating organic matter in aquatic systems with the help of an inadequate proxy will not only prevent us from obtaining a better understanding of how far organic matter is transported and when it outgasses into the atmosphere, it will also generate misleading conclusions about is stability and reactivity[3]. As outlined above for the soil environment, we argue that the persistence and movement of terrestrially derived organic carbon compounds entering aquatic ecosystems will rely on their protection by minerals, solubility in water and microbial degradation rather than primarily their chemical properties.

Aquatic carbon is not only important as part of the global carbon cycle, but also for local biogeochemical processes in streams and lakes. The observation of electron shuttling by 'humic substances' may serve as an example[77–79]. Electron shuttling is often attributed to quinones[80] and is a key driver for the microbial use of organic carbon, including organic pollutants and oxidation of reduced metals in oxygen-limited environments such as aquatic sediments and peatlands[81]. Extracts of 'humic substances' typically used for investigations of electron shuttling phenomena may have developed this capacity not as a result of 'humification', but because alkaline solutions extract quinones that are present in soil as a result of known microbial metabolism[82] or in carbon thermally altered by fire[83], which has been shown to be electrochemically active[84,85]. Abandoning the 'humic' proxy will broaden future research to include electron transfer mediated by organic matter that is not soluble in alkali. This will improve identification of mechanisms controlling methane production in temporarily anoxic environments[79] and those elements of biotic[85] and abiotic[86] iron cycles that remain elusive.

Water treatment is a vital technology, but its mechanistic basis is rendered questionable by the pervasive use of the 'humic substances' proxy. Anaerobic bioremediation refers to 'humic substances' as an electron acceptor[78] that removes pollutants. During purification of drinking water, on the other hand, 'humic acids' are considered contaminants, because reactions with disinfectants generate by-products that are toxic to humans[87]. Research specifically targets 'humic' isolates that are perceived to be relevant proxies for organic compounds in waste water[88]. Instead, water treatment would benefit from using organic materials that are based on mixtures of actually existing degradation products rather than the proxies based on alkaline extraction, as in the removal of organic matter by coagulation[87]. Water treatment needs to become more predictable because future contamination will inevitably include new pharmaceuticals or nanoparticles of which we have limited experience.

Agriculture

Productive soils are central to human welfare because agriculture generates most of our food, feed and fibre. Organic matter contributes to soil fertility by retaining plant-available water and nutrients or promoting the formation of soil structure, but it is also consumed in the process of arable soil management as it releases needed nutrients and energy when it decomposes[89]. However, proposals to return the carbon lost through agricultural activities in previous decades often emphasize the need to build or augment a 'stable humus' pool, drawing on the outdated concept of 'humification'. Such a pool has been suggested to increase soil organic matter resistance to decomposition through *in situ* synthesis of macromolecules[90] or hydrophobic protection by 'humic substances'[91]. However, this goal seems counterproductive given that soil organic matter is most beneficial when it decays and releases energy and nutrients[89]. Acknowledging the dynamic continuum of decomposition products suggests that the management of soil organic matter turnover is more important than the accrual of non-productive organic matter deposits. This requires a mechanistic understanding of interactions with minerals, movement into areas of lower mineralization and mediation of microbial activity[7]. The need to manage the turnover and volume of organic compounds and nutrient provisioning to optimize soil productivity (Fig. 1) warrants further research into balancing both stocks and flows of organic matter.

Soil organic matter can reduce contaminant uptake into crops and leaching into groundwater through adsorption at the cost of long-term accumulation. Studying the hypothetical interactions of heavy metals or other pollutants with extracts of 'humic substances' will provide limited insight into contaminant behaviour. Future research into interactions of organic matter with arsenic[92], other heavy metals[93] or pharmaceuticals[94] will generate more robust information by investigating the entire soil organic matter or the portion present in soil solution rather than what is extractable by alkali. This will allow better predictions of contaminant movement and mitigation of their environmental impact by adsorption and microbial use.

Alkaline extraction targets materials with abundant functional groups. Consequently, plant growth is often enhanced when such materials are added to soils particularly to stimulate rooting[95]. Alkali-extracted products are therefore becoming increasingly popular as soil amendments[96]. Better crop nutrition is an important part of this strategy and plant uptake of micronutrients is indeed known to be improved when organic compounds make them more soluble[97]. Positive plant responses to 'humic substances' resembling those of beneficial plant hormones[95], through improved defence mechanisms against pests or diseases[98] and changes in gene expression[99] may mean that the alkaline extracts contain compounds that trigger these effects. If we acknowledge soil organic matter as a continuum of decomposition products, we will be better able to design soil applications for specific purposes such as improved plant defence, and unpack what is essentially a 'black box' of compounds extracted by alkali. Research and product development should therefore focus on organic compounds that are soluble in water for managing soil health and focus on relationships between specific functional groups or compounds and positive plant responses for which information already exists.

The way forward

The need for the soil sciences to move away from both the 'humification' model and associated 'humic' language has been much debated. Unfortunately, this objective has not been implemented with rigour and has largely been ignored in the neighbouring fields of aquatic and environmental sciences. In many cases, the 'humification' model itself has been abandoned, but the 'humic' nomenclature is maintained. For example, the large molecular size of 'humic substances' has been refuted[13,100] but not their existence. The issue has also been approached by redefining 'humic substances' as the portion of soil organic matter that cannot be molecularly characterized[20,101,102], or by calling all soil organic matter 'humus'[11]. We argue that this compromise—maintaining terminology but altering its meanings in varying ways—hampers scientific progress beyond the soil sciences. The SCM of soil organic matter does not allow a confusing middle path; it requires leaving the traditional view behind to bring about lasting innovation and progress[103]. This is critical as scientific fields outside the soil sciences base their research on the false premise of the existence of 'humic substances'. Thus an issue of terminology becomes a problem of false inference, with far-reaching implications beyond our ability to communicate scientifically accurate soil processes and properties.

Reconciliation of modern experimental evidence with a robust molecular model can immediately be achieved by consistently referring to 'humic substances' as alkaline extracts rather than suggesting that

a distinct category of organic materials exists. This is essential when modelling global soil carbon, for which we need to cease using soil carbon pools whose definitions are rooted in 'humic' theory. In future research, alkaline extracts should not be used as proxies for naturally occurring organic matter or a subset thereof. Alkaline extraction should be supplanted by approaches that capture actual solubility in soil, river or ocean water.

The SCM will direct fundamental research questions towards microbial access to 'protected' rather than 'stable' carbon, and this will lead to more mechanistic representations of pollutant mobility and electron transfer reactions. In applied science and industry, this shift will prove more difficult to establish, because commercial 'humification' products and their marketing are strongly established, particularly in the gardening and compost industry. However, alkaline extraction does indeed isolate organic materials rich in oxygen, which may have value for product development. Therefore, we urgently need a biologically based explanation of the established growth-promoting effects of some highly oxidized organic compounds in soil in order to develop commercial products that operate in a predictable manner based on observable reactions of enzymes, hormones or cell wall transport. This will redirect existing research and development programmes at the intersection of molecular biology, ecology and soil biogeochemistry to allow the implementation of scientifically sound 'soil health' concepts.

Government-funded research programmes must therefore preferentially support science that bridges the gap between detailed and fine-scale mechanistic research at the plant–soil interface and field-scale research relevant to those who manage soils for their multiple ecosystem services. There are great opportunities for progress in explaining soil carbon responses to warming, and in the improvement of soil fertility and water quality. Coordinated interdisciplinary research programmes should be urgently set up to encourage greater coordination between soil biogeochemists and modellers. Such programmes should use the SCM to examine the balance between managing carbon and nutrient flows with sequestration, and between carbon transport, deposition and evasion in rivers and oceans. Models based on pools should be replaced with models based on organic matter solubility and spatial architecture to improve climate prediction, regional and global assessments of soil resources and soil vulnerability. The reward will be more robust forecasts and resource evaluation, issues critical for developing future climate change and land use policies.

Received 30 April 2015; accepted 8 October 2015.
Published online 23 November 2015.

1. Ciais, P. *et al.* in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) 465–570 (Cambridge Univ. Press, 2013).
2. Lal, R. Soil carbon sequestration impacts on global climate change and food security. *Science* **304,** 1623–1627 (2004).
3. Marín-Spiotta, E. *et al.* Paradigm shifts in soil organic matter research affect interpretations of aquatic carbon cycling: transcending disciplinary and ecosystem boundaries. *Biogeochemistry* **117,** 279–297 (2014).
4. Brady, N. C. & Weil, R. R. *The Nature and Properties of Soils* 14th edn, Ch. 12 (Prentice Hall, 2008).
5. Tan, K. H. *Humic Matter in Soil and the Environment: Principles and Controversies* 2nd edn (CRC Press, 2014).
6. Horwath, W. R. in *Soil Microbiology, Ecology, and Biochemistry* 4th edn (ed. Paul, E.) 339–382 (Academic Press, 2015).
7. Trumbore, S. E. Potential responses of soil organic carbon to global environmental change. *Proc. Natl Acad. Sci. USA* **94,** 8284–8291 (1997).
8. Schnitzer, M. & Monreal, C. M. Quo Vadis soil organic matter research? A biological link to the chemistry of humification. *Adv. Agron.* **113,** 143–217 (2011).
9. Achard, F. K. Chemische Untersuchung des Torfs. *Chem. Ann. Freunde Naturlehre, Arzneigel. Haushalt. Manufact.* **2,** 391–403 (1786).
10. van Bemmelen, J. M. Die Absorptionsverbindungen und das Absorptionsvermögen der Ackererde. *Landwirtschaftlichen Versuchs-Stationen* **35,** 69–136 (1888).
11. Waksman, S. A. *Humus. Origin, Chemical Composition and Importance in Nature* (Williams and Wilkins, 1936).
This is the first major critique of the humification concept.
12. Stevenson, F. J. *Humus Chemistry: Genesis, Composition, Reactions* (John Wiley & Sons, 1994).
13. Sutton, R. & Sposito, G. Molecular structure in soil humic substances: the new view. *Environ. Sci. Technol.* **39,** 9009–9015 (2005).
14. Stumm, W. & Morgan, J. J. *Aquatic Chemistry, An Introduction Emphasizing Chemical Equilibria in Natural Waters* (Wiley, 1981).
15. Burdon, J. Are the traditional concepts of the structures of humic substances realistic? *Soil Sci.* **166,** 752–769 (2001).
16. Jenkinson, D. S. & Rayner, J. H. The turnover of soil organic matter in some of the Rothamsted classical experiments. *Soil Sci.* **123,** 298–305 (1977).
17. Weiss, M. S. *et al.* Molecular architecture and electrostatic properties of a bacterial porin. *Science* **254,** 1627–1630 (1991).
18. Hedges, J. I. & Oades, J. M. Comparative organic geochemistries of soils and marine sediments. *Org. Geochem.* **27,** 319–361 (1997).
19. Baldock, J. A. & Nelson, P. N. in *Handbook of Soil Science* (ed. Sumner, M. E.) B25–B84 (CRC Press, 2000).
20. Essington, M. E. *Soil and Water Chemistry: An Integrated Approach* (CRC Press, 2004).
21. Bosatta, E. & Agren, G. I. Dynamics of carbon and nitrogen in the organic-matter of the soil—a generic theory. *Am. Nat.* **138,** 227–245 (1991).
22. Kononova, M. M. *Soil Organic Matter. Its Nature, its Role in Soil Formation and in Soil Fertility* (Pergamon, 1961).
23. Hayes, M. H. B. & Swift, R. S. in *Soil Colloids and their Associations in Aggregates* (eds De Boodt, M. F., Hayes, M. H. B. & Herbillon, A.) 245–305 (Plenum, 1990).
24. Guggenberger, G. in *Microorganisms in Soils: Roles in Genesis and Functions* (eds Buscot, F. & Varma, A.) 85–106 (Springer, 2005).
25. Sollins, P., Homann, P. & Caldwell, B. A. Stabilization and destabilization of soil organic matter: mechanisms and controls. *Geoderma* **74,** 65–105 (1996).
26. Melillo, J. M., Aber, J. D. & Muratore, J. F. Nitrogen and lignin control of hardwood leaf litter decomposition dynamics. *Ecology* **63,** 621–626 (1982).
27. Aber, J. D., Melillo, J. M. & McClaugherty, C. A. Predicting long-term patterns of mass-loss, nitrogen dynamics, and soil organic matter formation from initial litter chemistry in temperate forest ecosystems. *Can. J. Bot.* **68,** 2201–2208 (1990).
28. Baldock, J. A. *et al.* Aspects of the chemical structure of soil organic materials as revealed by solid-state ^{13}C NMR spectroscopy. *Biogeochemistry* **16,** 1–42 (1992).
29. Lützow, M. V. *et al.* Stabilization of organic matter in temperate soils: mechanisms and their relevance under different soil conditions—a review. *Eur. J. Soil Sci.* **57,** 426–445 (2006).
30. Gramss, G., Voigt, K. D. & Kirsche, B. Degradation of polycyclic aromatic hydrocarbons with three to seven aromatic rings by higher fungi in sterile and unsterile soils. *Biodegradation* **10,** 51–62 (1999).
31. Wiesenberg, G. L. B., Schwarzbauer, J., Schmidt, M. W. I. & Schwark, L. Source and turnover of organic matter in agricultural soils derived from *n*-alkane/*n*-carboxylic acid compositions and C-isotope signatures. *Org. Geochem.* **35,** 1371–1393 (2004).
32. Hamer, U., Marschner, B., Brodowski, S. & Amelung, W. Interactive priming of black carbon and glucose mineralisation. *Org. Geochem.* **35,** 823–830 (2004).
33. Hazen, T. C. *et al.* Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science* **330,** 204–208 (2010).
34. Yang, J., Yang, Y., Wu, W.-M., Zhao, J. & Jiang, L. Evidence of polyethylene biodegradation by bacterial strains from the guts of plastic-eating waxworms. *Environ. Sci. Technol.* **48,** 13776–13784 (2014).
These experiments show that under certain conditions bacteria can even utilize plastics, not thought to be biologically degradable.
35. Berg, B. & Staaf, H. Decomposition rate and chemical changes of Scots pine needle litter. II. Influence of chemical composition. *Ecol. Bull.* **32,** 373–390 (1980).
36. Klotzbücher, T., Kaiser, K., Guggenberger, G., Gatzek, C. & Kalbitz, K. A new conceptual model for the fate of lignin in decomposing plant litter. *Ecology* **92,** 1052–1062 (2011).
This paper shows that lignin in plant litter can be easily mineralized by microorganisms provided that bioavailable carbon is present.
37. Hedges, J. I. in *Humic Substances and their Role in the Environment* (eds Frimmel, F. H. & Christman, R. F.) 45–58 (John Wiley & Sons, 1988).
38. Cotrufo, M. F., Wallenstein, M. D., Boot, C. M., Denef, K. & Paul, E. The Microbial Efficiency-Matrix Stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: do labile plant inputs form stable soil organic matter? *Glob. Change Biol.* **19,** 988–995 (2013).
This paper discusses the need to integrate concepts of litter decomposition with protection and mineralization of soil organic matter, and proposes a solution.
39. Kelleher, B. P. & Simpson, A. J. Humic substances in soils: are they really chemically distinct? *Environ. Sci. Technol.* **40,** 4605–4611 (2006).
This paper provides direct proof that the chemical makeup of 'humic substances' can be explained as resulting from mixtures of known plant and microbial compounds.
40. Gillespie, A. *et al.* Glomalin-related soil protein contains non-mycorrhizal-related heat-stable proteins, lipids and humic materials. *Soil Biol. Biochem.* **43,** 766–777 (2011).
41. Mylotte, R. *et al.* Isolation and characterisation of recalcitrant organic components from an estuarine sediment core. *J. Soils Sedim.* **15,** 211–224 (2015).
42. Lehmann, J. *et al.* Spatial complexity of soil organic matter forms at nanometre scales. *Nature Geosci.* **1,** 238–242 (2008).

This paper demonstrates that spectral properties of alkaline extracts are not found in soils at a high spatial resolution.

43. Schumacher, M., Christl, I., Scheinost, A. C., Jacobsen, C. & Kretzschmer, R. Chemical heterogeneity of organic soil colloids investigated by scanning transmission X-ray microscopy and C-1s NEXAFS spectroscopy. *Environ. Sci. Technol.* **39,** 9094–9100 (2005).
44. Lehmann, J., Kinyangi, J. & Solomon, D. Organic matter stabilization in soil microaggregates: implications from spatial heterogeneity of organic carbon contents and carbon forms. *Biogeochemistry* **85,** 45–57 (2007).
45. Chen, C., Dynes, J. J., Wang, J., Karunakaran, C. & Sparks, D. L. Soft X-ray spectromicroscopy study of mineral-organic matter associations in pasture soil clay fractions. *Environ. Sci. Technol.* **48,** 6678–6686 (2014).
46. Myneni, S. C. B., Brown, J. T., Martinez, G. A. & Meyer-Ilsel, W. Imaging of humic substance macromolecular structures in water and soils. *Science* **286,** 1335–1337 (1999).
This paper provides spectroscopic evidence that organic materials in alkaline extracts are assemblages of smaller compounds mimicking larger molecules.
47. Nebbioso, A. & Piccolo, A. Advances in humeomics: enhanced structural identification of humic molecules after size fractionation of a soil humic acid. *Anal. Chim. Acta* **720,** 77–90 (2012).
48. Hedges, J. I. *et al.* The molecularly-uncharacterized component of nonliving organic matter in natural environments. *Org. Geochem.* **31,** 945–958 (2000).
49. Miltner, A., Bombach, P., Schmidt-Brücken, B. & Kästner, M. SOM genesis: microbial biomass as a significant source. *Biogeochemistry* **111,** 41–55 (2012).
50. Schurig, C. *et al.* Microbial cell-envelope fragments and the formation of soil organic matter: a case study from a glacier forefield. *Biogeochemistry* **113,** 595–612 (2013).
51. Hedges, J. I. & Keil, R. G. Organic geochemical perspectives on estuarine processes: sorption reactions and consequences. *Mar. Chem.* **65,** 55–65 (1999).
52. Schmidt, M. W. I. *et al.* Persistence of soil organic matter as an ecosystem property. *Nature* **478,** 49–56 (2011).
53. Stockmann, U. *et al.* The knowns, known unknowns and unknowns of sequestration of soil organic carbon. *Agric. Ecosyst. Environ.* **164,** 80–99 (2013).
54. Parton, W. J. *et al.* ForCent model development and testing using the Enriched Background Isotope Study experiment. *J. Geophys. Res. Biogeosci*. **115,** G4 (2010).
55. Jenkinson, D. S. & Coleman, K. The turnover of organic carbon in subsoils. Part 2. Modelling carbon turnover. *Eur. J. Soil Sci.* **59,** 400–413 (2008).
56. Todd-Brown, K. E. O. *et al.* Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeoscience* **10,** 1717–1736 (2013).
57. Bonan, G. B., Hartman, M. D., Parton, W. J. & Wieder, W. R. Evaluating litter decomposition in earth system models with long-term litterbag experiments: an example using the Community Land Model version 4 (CLM4). *Glob. Change Biol.* **19,** 957–974 (2013).
58. Koven, C. D. *et al.* The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. *Biogeoscience* **10,** 7109–7131 (2013).
59. Davidson, E. A. & Janssens, I. A. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* **440,** 165–173 (2006).
60. Conant, R. T. *et al.* Sensitivity of organic matter decomposition to warming varies with its quality. *Glob. Change Biol.* **14,** 868–877 (2008).
61. Craine, J. M., Fierer, N. & McLauchlan, K. K. Widespread coupling between the rate and temperature sensitivity of organic matter decay. *Nature Geosci*. **3,** 854–857 (2010).
62. Bosatta, E. & Agren, G. I. Soil organic matter quality interpreted thermodynamically. *Soil Biol. Biochem.* **31,** 1889–1891 (1999).
63. Arrhenius, S. Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren. *Z. Phys. Chem.* **4,** 226–248 (1889).
64. Kögel-Knabner, I. The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. *Soil Biol. Biochem.* **34,** 139–162 (2002).
65. Erhagen, B. *et al.* Temperature response of litter and soil organic matter decomposition is determined by chemical composition of organic material. *Glob. Change Biol.* **19,** 3858–3871 (2013).
66. Sierra, C. A., Trumbore, S. E., Davidson, E. A., Vicca, S. & Janssens, I. Sensitivity 0of decomposition rates of soil organic matter with respect to simultaneous changes in temperature and moisture. *J. Adv. Modeling Earth Syst.* **7,** 335–356 (2015).
67. Kleber, M. *et al.* Mineral–organic associations: formation, properties, and relevance in soil environments. *Adv. Agron.* **130,** 1–140 (2015).
This is a comprehensive discussion of the role of the mineral matrix in controlling organic matter transformations and persistence.
68. Keiluweit, M. *et al.* Mineral protection of soil carbon counteracted by root exudates. *Nature Clim. Change* **5,** 588–595 (2015).
69. Todd-Brown, K. E., Hopkins, F. M., Kivlin, S. N., Talbot, J. M. & Allison, S. D. A framework for representing microbial decomposition in coupled climate models. *Biogeochemistry* **109,** 19–33 (2012).
70. Treseder, K. K. *et al.* Integrating microbial ecology into ecosystem models: challenges and priorities. *Biogeochemistry* **109,** 7–18 (2012).
71. Nunan, N. *et al.* Metabolising old soil carbon: simply a matter of simple organic matter? *Soil Biol. Biochem.* **88,** 128–136 (2015).
72. Riley, W. J. *et al.* Long residence times of rapidly decomposable soil organic matter: application of a multi-phase, multi-component, and vertically resolved model (BAMS1) to soil carbon dynamics. *Geosci. Model Dev*. **7,** 1335–1355 (2014).
73. Dal Ferro, N. & Morari, F. From real soils to 3D-printed soils: reproduction of complex pore network at the real size in a silty-loam soil. *Soil Sci. Soc. Am. J.* **79,** 1008–1017 (2015).
74. Richey, J. E. *et al.* Outgassing from the Amazonian rivers and wetlands as a large tropical source of atmospheric CO_2. *Nature* **416,** 617–620 (2002).
75. Cole, J. J. *et al.* Plumbing the global carbon cycle: integrating inlands waters into the terrestrial carbon budget. *Ecosystems* **10,** 172–185 (2007).
76. Hedges, J. I. & Keil, R. G. Sedimentary organic matter preservation: an assessment and speculative synthesis. *Mar. Chem.* **49,** 81–115 (1995).
77. Lovley, D. R., Coates, J. D., Blunt-Harris, E. L., Phillips, E. J. & Woodward, J. C. Humic substances as electron acceptors for microbial respiration. *Nature* **382,** 445–448 (1996).
78. Martinez, C. M., Alvarez, L. H., Celis, L. B. & Cervantes, F. J. Humus-reducing microorganisms and their valuable contribution in environmental processes. *Appl. Microbiol. Biotechnol.* **97,** 10293–10308 (2013).
79. Klüpfel, L., Piepenbrock, A., Kappler, A. & Sander, M. Humic substances as fully regenerable electron acceptors in recurrently anoxic environments. *Nature Geosci*. **7,** 195–200 (2014).
80. Uchimiya, M. & Stone, A. T. Reduction of substituted *p*-benzoquinones by Fe^{II} near neutral pH. *Aquat. Geochem.* **16,** 173–188 (2010).
81. Keller, J. K. & Takagi, K. K. Solid-phase organic matter reduction regulates anaerobic decomposition in bog soils. *Ecosphere* **4,** 54 (2013).
82. Newman, D. K. & Kolter, R. A role for excreted quinones in extracellular electron transfer. *Nature* **405,** 94–97 (2001).
83. Solomon, D. *et al.* Molecular signature and sources of biochemical recalcitrance of organic C in Amazonian Dark Earths. *Geochim. Cosmochim. Acta* **71,** 2285–2298 (2007).
84. Klüpfel, L., Keiluweit, M., Kleber, M. & Sander, M. Redox properties of plant biomass-derived black carbon (biochar). *Environ. Sci. Technol.* **48,** 5601–5611 (2014).
85. Kappler, A. *et al.* Biochar as an electron shuttle between bacteria and Fe(III) minerals. *Environ. Sci. Technol. Lett.* **1,** 339–344 (2014).
86. Melton, E. D., Swanner, E. D., Behrens, S., Schmidt, C. & Kappler, A. The interplay of microbially mediated and abiotic reactions in the biogeochemical Fe cycle. *Nature Rev. Microbiol.* **12,** 797–808 (2014).
87. Wang, W. *et al.* Effects of UV radiation on humic acid coagulation characteristics in drinking water treatment processes. *Chem. Eng. J.* **256,** 137–143 (2014).
88. Richardson, S. D. & Postigo, C. in *Emerging Organic Contaminants and Human Health* (ed. Barceló, D.) 93–137 (Springer, 2012).
89. Janzen, H. H. Beyond carbon sequestration: soil as conduit of solar energy. *Eur. J. Soil Sci.* **66,** 19–32 (2015).
90. Lorenz, K., Lal, R., Preston, C. M. & Nierop, K. G. J. Strengthening the soil organic carbon pool by increasing contributions from recalcitrant aliphatic bio(macro)molecules. *Geoderma* **142,** 1–10 (2007).
91. Spaccini, R., Piccolo, A., Conte, P., Haberhauer, G. & Gerzabek, M. H. Increased soil organic carbon sequestration through hydrophobic protection by humic substances. *Soil Biol. Biochem.* **34,** 1839–1851 (2002).
92. Fakour, H. & Lin, T. F. Experimental determination and modeling of arsenic complexation with humic and fulvic acids. *J. Hazard. Mater.* **279,** 569–578 (2014).
93. Tang, W. W. *et al.* Impact of humic/fulvic acid on the removal of heavy metals from aqueous solutions using nanomaterials: a review. *Sci. Total Environ.* **468–469,** 1014–1027 (2014).
94. Aristilde, L. & Sposito, G. Complexes of the antimicrobial ciprofloxacin with soil, peat, and aquatic humic substances. *Environ. Toxicol. Chem.* **32,** 1467–1478 (2013).
95. Calvo, P., Nelson, L. & Kloepper, J. W. Agricultural uses of plant biostimulants. *Plant Soil* **383,** 3–41 (2014).
96. Rose, M. T. *et al.* A meta-analysis and review of plant-growth response to humic substances: practical implications for agriculture. *Adv. Agron.* **124,** 37–89 (2014).
97. Bocanegra, M. P., Lobartini, J. C. & Orioli, G. A. Plant uptake of iron chelated by humic acids of different molecular weights. *Commun. Soil Sci. Plant Anal.* **37,** 239–248 (2006).
98. Berbara, R. L. L. & García, A. C. in *Physiological Mechanisms and Adaptation Strategies in Plants under Changing Environment* (eds Ahmad, P. & Wani, M. R.) 297–319 (Springer, 2014).
99. Trevisan, S. *et al.* Humic substances affect *Arabidopsis* physiology by altering the expression of genes involved in primary metabolism, growth and development. *Environ. Exp. Bot.* **74,** 45–55 (2011).
100. Wershaw, R. L. *Evaluation of Conceptual Models of Natural Organic Matter (Humus) From a Consideration of the Chemical and Biochemical Processes of Humification*. US Geological Survey Scientific Investigations Report 2004–5121, http://pubs.usgs.gov/sir/2004/5121/pdf/sir2004-5121.pdf (USGS, 2004).
101. Baldock, J. A. & Broos, K. in *Handbook of Soil Sciences: Resource Management and Environmental Impacts* (eds Huang, P. M., Li, Y. & Sumner, M. E.) 11-1–52 (CRC Press, 2011).

102. Hatcher, P. G. The CHNs of organic geochemistry: characterization of molecularly uncharacterized non-living organic matter. *Mar. Chem.* **92,** 5–8 (2004).
103. Kuhn, T. S. *The Structure of Scientific Revolutions* (Chicago Press, 1962).
104. Forbes, M. S., Raison, R. J. & Skjemstad, J. O. Formation, transformation and transport of black carbon (charcoal) in terrestrial and aquatic ecosystems. *Sci. Total Environ.* **370,** 190–206 (2006).
105. International Humic Substances Society (IHSS). *What are Humic Substances?* http://www.humicsubstances.org/whatarehs.html (IHSS, 2015).
106. Rice, J. A. Humin. *Soil Sci.* **166,** 848–857 (2001).
107. Tatzber, M. *et al.* Decomposition of carbon-14-labeled organic amendments and humic acids in a long-term field experiment. *Soil Sci. Soc. Am. J.* **73,** 744–750 (2009).
108. Gramss, G., Ziegenhagen, D. & Sorge, S. Degradation of soil humic extract by wood- and soil-associated fungi, bacteria, and commercial enzymes. *Microb. Ecol.* **37,** 140–151 (1999).
This paper provides experimental evidence for rapid metabolization of 'humic extracts' by a selection of soil fungi.
109. Maillard, L. C. Synthèse des matières humiques par action des acides aminés sur les sucres réducteurs. *Ann. Chimie* **5,** 258–317 (1916).
110. Huang, P. M. & Hardie, A. G. in *Biophysico-Chemical Processes Involving Natural Nonliving Organic Matter in Environmental Systems* (eds Senesi, N., Xing, B. & Huang, P. M.) 41–109 (John Wiley & Sons, 2009).
111. Piccolo, A. The supramolecular structure of humic substances. *Soil Sci.* **166,** 810–832 (2001).
112. Campbell, C. A., Paul, E. A., Rennie, D. A. & McCallum, K. J. Applicability of the carbon-dating method of analysis to soil humus studies. *Soil Sci.* **104,** 217–224 (1967).
113. Gleixner, G. Soil organic matter dynamics: a biological perspective derived from the use of compound-specific isotopes studies. *Ecol. Res.* **28,** 683–695 (2013).
114. Chen, J., Henderson, G., Grimm, C. C., Lloyd, S. W. & Laine, R. A. Termites fumigate their nests with naphthalene. *Nature* **392,** 558–559 (1998).
115. Staunton, J. & Weissmann, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18,** 380–416 (2001).
116. Skjemstad, J. O., Clarke, P., Taylor, J. A., Oades, J. M. & McClure, S. G. The chemistry and nature of protected carbon in soil. *Aust. J. Soil Res.* **34,** 251–271 (1996).
117. Lehmann, J. *et al.* Australian climate–carbon cycle feedback reduced by soil black carbon. *Nature Geosci.* **1,** 832–835 (2008).
118. Rodionov, A. *et al.* Black carbon in grassland ecosystems of the world. *Glob. Biogeochem. Cycles* **24,** GB3013 (2010).
119. Heymann, K. *et al.* Can functional group composition of alkaline isolates from black carbon-rich soils be identified on a sub-100 nm scale? *Geoderma* **235–236,** 163–169 (2014).
120. Knicker, H. How does fire affect the nature and stability of soil organic nitrogen and carbon? A review. *Biogeochemistry* **85,** 91–118 (2007).
121. Knicker, H., Hilscher, A., Gonzalez-Vila, F. J. & Almendros, G. A new conceptual model for the structural properties of char produced during vegetation fires. *Org. Geochem.* **39,** 935–939 (2008).
122. LeBoeuf, E. J. & Weber, W. J. Macromolecular characteristics of natural organic matter. 1. Insights from glass transition and enthalpic relaxation behavior. *Environ. Sci. Technol.* **34,** 3623–3631 (2000).
123. Schaumann, G. E. Soil organic matter beyond molecular structure. Part I: macromolecular and supramolecular characteristics. *J. Plant Nutr. Soil Sci.* **169,** 145–156 (2006).
124. Chilom, G. & Rice, J. A. Glass transition and crystallite melting in natural organic matter. *Org. Geochem.* **36,** 1339–1346 (2005).
125. Keiluweit, M., Nico, P. S., Johnson, M. G. & Kleber, M. Dynamic molecular structure of plant biomass-derived black carbon (biochar). *Environ. Sci. Technol.* **44,** 1247–1253 (2010).

Author Contributions Both authors contributed equally to the concept, outline, and writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.L. (CL273@cornell.edu).

doi:10.1038/nature15744

Soil biodiversity and human health

Diana H. Wall[1]*, Uffe N. Nielsen[2]* & Johan Six[3]*

Soil biodiversity is increasingly recognized as providing benefits to human health because it can suppress disease-causing soil organisms and provide clean air, water and food. Poor land-management practices and environmental change are, however, affecting belowground communities globally, and the resulting declines in soil biodiversity reduce and impair these benefits. Importantly, current research indicates that soil biodiversity can be maintained and partially restored if managed sustainably. Promoting the ecological complexity and robustness of soil biodiversity through improved management practices represents an underutilized resource with the ability to improve human health.

Soils comprise a dynamic reservoir of biodiversity within which the interactions between microbes, animals and plants provide many benefits for human well-being; however, their potential use for the maintenance of human health has been less clear[1–3]. Living soils are vital to humans because soil biodiversity, with its inherent complexity (the types, sizes, traits and functions of soil organisms), not only provides disease control but also influences the quantity and quality of the food we eat, the air we breathe and the water we drink[4]. The long-term provision of these benefits is dependent on careful and sustainable use of soils as a resource. Yet, soil biodiversity is often unintentionally affected by human-induced global changes. Land-use change, including urbanization, agriculture, deforestation and desertification, can have a ripple effect on soils and soil biodiversity that extends far beyond the original site of disturbance[5,6]. For example, the increase in soil erosion by water and wind contributes to the formation of dust storms and the dispersal of soil organisms and pathogens, with effects on soil biodiversity and ultimately on human, plant and animal health[7–10].

Research efforts are rapidly producing information about soil biodiversity and its functions, which can be combined with land managers' knowledge, to inform the development of sustainable soil-management practices[1,11–13]. The resulting global preservation and restoration of soils would provide an additional path towards decreasing disease in and providing medicine for humans, plants and animals.

Here, we focus on the impacts of the use and mismanagement of land on human health due to (1) changes in the prevalence of antagonists for soil-borne pests and pathogens that cause diseases in humans, plants and animals, and (2) changes in soil biodiversity that affect the maintenance of health (Fig. 1). We use the integrated concept of human

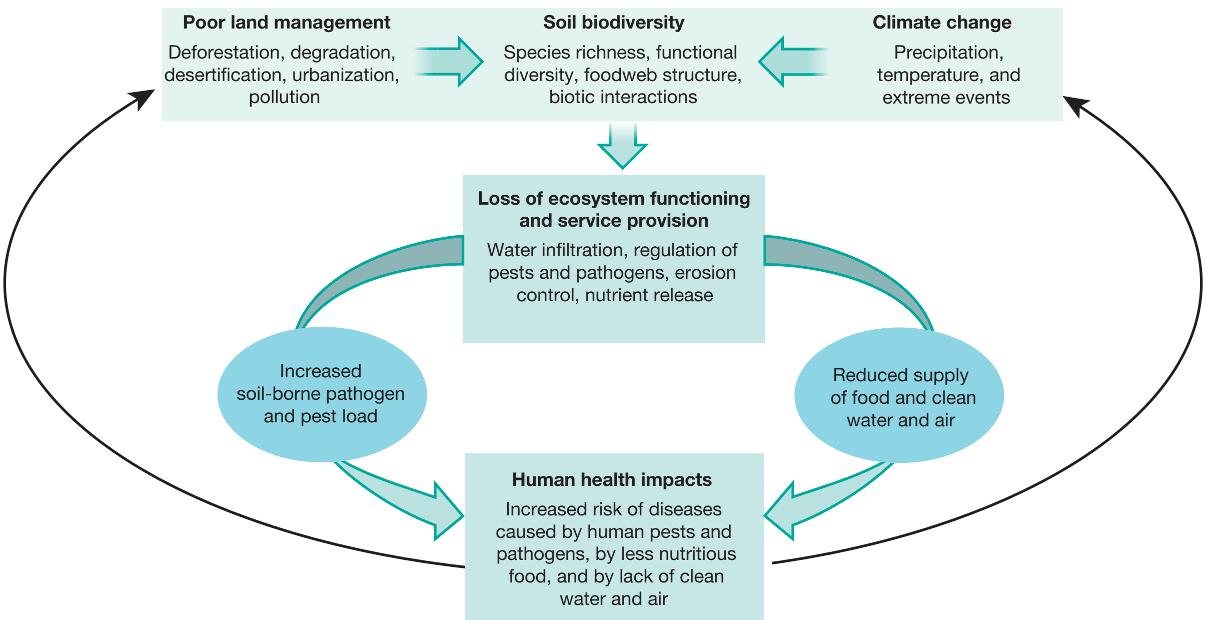


Figure 1 | Flow diagram illustrating the link between soil biodiversity and human health. Soil biodiversity is often negatively affected by the interaction between poor land management practices and drivers of climate change, both of which ultimately compromise ecosystem function and services that are essential for human health (control of pests and pathogens, production of nutritious food, cleansing water and reducing air pollution). Responses to reduced human health can in turn affect management decisions that govern land use and climate change.

[1]School of Global Environmental Sustainability and Department of Biology, Colorado State University, Fort Collins, Colorado 80523-1036, USA. [2]Hawkesbury Institute for the Environment, Locked Bag 1797, Western Sydney University, Penrith, New South Wales 2751, Australia. [3]Department of Environmental Systems Science, Swiss Federal Institute of Technology ETH-Zurich, Zurich 8092, Switzerland.
*These authors contributed equally to this work.

Table 1 | Soil pathogens and parasites of humans, animals and plants

		Euedaphic pathogens		Soil-transmitted pathogens	
Host organism	Type of pathogen	Species name	Disease caused	Species name	Disease caused
Humans	Bacteria	*Bacillus anthracis* *Listeria monocytogenes*	Anthrax Listeriosis	*Escherichia coli* *Salmonella* spp.	Diarrhoea Salmonellosis, Typhoid fever, Diarrhoea
	Fungi	*Aspergillus* spp. *Coccidioides immitis* *Histoplasma capsulatum*	Aspergillosis Valley fever Histoplasmosis		
	Protozoa	*Naegleria fowleri*	Brain encephalitis	*Toxoplasma gondii*	Toxoplasmosis
	Helminths (Nematoda)			*Ascaris lumbricoides* *Ancylostoma duodenale* *Necator americanus* *Strongyloides stercoralis*	Ascariasis Hookworm Hookworm Strongyloidiasis
	Platyhelminthes			*Taenia saginata*	Beef tapeworm
Animals	Bacteria	*Bacillus anthracis*	Anthrax		
	Helminths (Nematoda)			*Haemonchus contortus*	Haemonchosis
Plants	Bacteria	*Agrobacterium tumefaciens*	Crown gall		
	Fungi	*Phytophthora infestans*	Potato blight		
	Helminths (Nematoda)	*Meloidogyne* spp. *Bursaphelenchus xylophilus*	Root knot Pine wood		

Following refs 19 and 89, pathogens are listed as euedaphic (true soil organisms) or as soil-transmitted (those temporarily living in soil and transmitted to a host).

health as defined by The World Health Organization and Convention on Biological Diversity[14], which extends beyond disease and infirmity and recognizes human connections to other species, ecosystems and the ecological foundation of varied drivers and protectors of human health. We specifically discuss how knowledge of the linkages between soil biodiversity and human health can be strengthened for improved management of land. Some aspects of land-use change in relation to soil biodiversity and human health are covered elsewhere, including industrial pollution, radioactivity, landfills, resource extraction, and mineral toxicity; and are not included here[15,16].

Soil biodiversity and soil-borne pathogens

Most soil organisms pose no risk to human health; rather, evidence is accumulating that soil biodiversity can be of great benefit[17,18]. Soil-borne pathogens and parasites that cause human diseases represent a minority of the species living in soils. There is a great opportunity to capitalize on the positive effects of soil organisms on human health through their roles (direct and indirect) in controlling soil-borne pathogens and pests (listed in Table 1).

Many animal, plant and human disease-causing organisms or their vectors live in soil, but their relationship to human diseases and the environment is not fully elucidated[16,19–21]. To address soil management and public health we need an understanding of the organisms, their ecological interactions, and why they become prevalent or persistent in soils[22,23]. Some soil-borne pathogens, such as the bacterial genera *Pseudomonas* and *Enterobacter*, are opportunistic species that can infect and cause diseases in humans but whose main functions in the soil foodweb are as antagonists against plant root pathogens, promoters of plant growth and decomposers[20,24]. Other soil-borne pathogens are obligate parasites that require a host to complete their life cycle. Most of these organisms can survive in soils for weeks to years, including as spores and eggs or inside carcasses. Soil-borne pathogens causing human infectious disease can be either true inhabitants of soils (euedaphic) or are transmitted via soils (Table 1). Soil-transmitted pathogens are usually obligate pathogens and reside temporarily in soil before being transmitted to humans by contact, vectors or in faeces.

Soil and anthrax

Anthrax is a zoonotic disease infecting humans, wildlife and livestock caused by the bacterium *Bacillus anthracis*. Known in the USA as an agent of bioterrorism, *B. anthracis* is relatively common and found in soils worldwide, including within the USA. Anthrax spores can remain dormant in soils for decades, but with heavy rains they are brought to the soil surface and attach to roots and grasses, which are grazed by animals. There have been outbreaks in eastern Colorado and Texas, occasionally resulting in die-offs of grazing animals, usually cattle. In contrast to this episodic occurrence, *B. anthracis* in Namibia and east Africa occurs annually in zebras and other grazing animals. In recent field experiments in Namibia, carcasses of animals infected with *B. anthracis* were shown to promote grass growth, which thus made the site more attractive for grazing wildlife[25]. This mechanism provides *B. anthracis* with a wildlife host and continues the cycle of the infectious disease.

Hence, one of the more effective strategies to reduce anthrax prevalence is burning of the vegetation, especially at sites of carcass deposition. However, maintaining soil cover is also important to reduce dust formation by wind erosion because human infections of anthrax typically result from inhalation of airborne spores or via vectors.

In general, soils favourable for anthrax are calcium-rich with neutral to alkaline pH (that is, Chernozem soils)[26]. Studies on soils and anthrax disease ecology in the Kruger National Park in South Africa found that when soil calcium was >150 milliequivalents per gram and pH >7, the anthrax death rate for ungulates was seven times higher than in other nearby soils[26].

Soil and helminths

The nematode *Strongyloides*, a soil-transmitted helminth and a parasite of humans and animals (Table 1), has a unique life cycle that alternates between free-living in soil and parasitic. The larvae are passed into soil in faeces and moult either (1) to become larvae that can infect humans or (2) to develop into adults that produce eggs and become a new free-living generation in soil. The free-living form feeds on bacteria as part of the soil foodweb, but its role in decomposition and nutrient cycling is not well understood. When infective larvae in soil come into contact with a suitable host, they penetrate the skin and eventually migrate to the intestine, where they reproduce. *Strongyloides stercoralis* infections occur in 10% to 40% of the human population in many tropical and subtropical countries[27], as a result of poor sanitation practices. In a study in rural Cambodia, about 45% of the people tested were infected, and a higher risk of infection was associated with lower organic carbon content of soils and land-use conversion from forest to cropland[28].

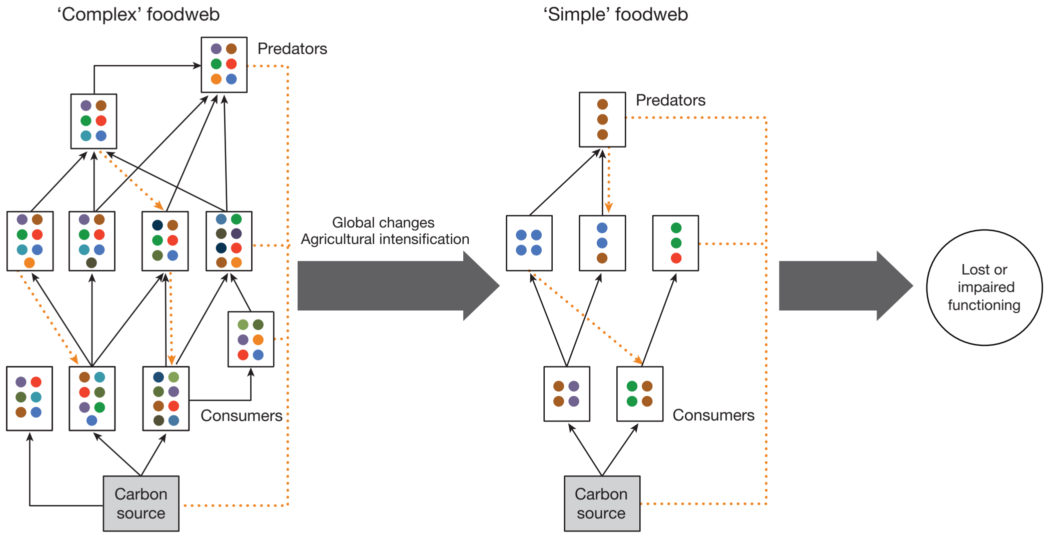
This strongly suggests that increasing the soil organic carbon levels in our croplands could be effective in reducing the prevalence of disease-causing helminths. Also included in the category of soil-transmitted helminths are hookworms and roundworms (Table 1), which infect many people globally. For example, in 2003, China and sub-Saharan Africa each had an estimated 200 million hookworm infections[29]. The contributions of hookworms and roundworms to the soil foodweb and their relationship to soil properties, however, are not well known.

To understand and predict the incidence of soil-borne pathogens and parasites in the future, much can be gained by integrative studies on their

BOX 1

Human influence on soil foodwebs

Soil foodwebs are naturally complex entities that enhance ecosystem functions such as biogeochemical cycling and the suppression of pests and pathogens for plants, animals and humans. Global changes, however, can have detrimental impacts on soil foodweb complexity by reducing belowground biodiversity and affecting biotic interactions. Substantial advances have been made in understanding these impacts on ecosystem functioning over the past few decades[1,3,65]. Furthermore, recent studies suggest that soil foodweb complexity is essential to maintaining high rates of ecosystem function[45,46,65,86]. Thus, activities that cause belowground biodiversity losses (such as loss of taxa and trophic levels) contribute to a reduction in foodweb complexity and, thereby, the capacity of soils to perform ecosystem functions. See Box 1 Figure, in which taxa are shown as coloured circles, trophic levels are shown as boxes, solid lines represent food sources and dashed lines indicate omnivory. The 'simple' foodweb on the right has been adversely affected by human-induced changes. Although these functions may not be lost completely, even reduced levels of functioning can influence human health directly (by reduced suppression of soil-borne diseases) or indirectly (by reduced provision of food, clean water and air). In some cases, these functions can be replaced through human interventions such as increased fertilizer and pesticide inputs, but promoting ecosystem functioning by managing soil biodiversity is likely to be more cost-effective and will ensure long-term sustainability.



Box 1 Figure | Key features in the reduction of species in soil food webs due to human influence and the adverse effect on ecosystem functioning.

life cycles in soils, their role in soil foodwebs, and how they are affected by environmental variables. The resulting knowledge can be used to establish viable management options to reduce the impacts of soil-borne pathogens and parasites. Integrating this new soil-based knowledge with the experience of public health researchers would provide an enormous opportunity for new soil-based approaches and policies to control current and emerging infectious diseases.

Soil and allergies

Several studies have shown that exposure to soil microorganisms lessens the prevalence of allergic diseases[22,30–33]. In particular, there is evidence that our immune system needs to be exposed to possible pathogens residing in soils in order to develop tolerance[34]. For example, it was found that individuals living in more urban environments have a lower diversity of bacteria on their skin and lower immunity expression[33,35]. It is predicted[14,33] that nearly two-thirds of the global human population will be living in urban areas by 2050 (refs 14, 33), resulting in less stimulation of our immune systems by soil organisms, and leading to more allergic diseases. Management of urban areas could easily consider access to natural areas and small livestock (chickens, ducks, rabbits and goats) as a way of exposing the urban population to soil organisms.

Soil, antibiotics and antihelminth resistance

As soils are altered through global change and associated losses in biodiversity above- and belowground, there is concern that we are losing a possible source of antibiotics and medicines, as well as the biological controls needed to prevent human, animal and plant disease. Antibiotic resistance to microbial-derived medicines has increased rapidly, threatening the prevention and treatment of diseases caused by bacteria, fungi and parasites[17]. The development of new antibiotics using soil has been very slow, because about 99% of bacteria have yet to be cultured. However, a new technique recently identified an antibiotic from an uncultured soil bacterium that can kill *Mycobacterium tuberculosis*, the causal agent of tuberculosis[17]. This is very promising: other as-yet-uncultured species may also reveal novel antibiotics. Helminth parasitic worm infections in humans, cattle and other domestic animals are often treated with antiparasitic, antihelminthic drugs, which are showing increasing resistance. Fortunately, land-management practices, including rotating pastures with more-resistant animals, breaking up or removing manure piles in pastures, or managing for higher grass growth so that animals do not graze on the parasites found in soils, are useful options for reducing the risk of parasite infection[36,37].

Soil and biological control

Human health is influenced indirectly by our choice of agricultural management practices owing to changes in the nutritional value of the plants and animals we eat, and the quantity of food produced. Plants are subject to many diseases caused by bacteria, fungi, viruses and parasites, which affect plant growth, nutrient levels and the quality of our food. In agriculture, biocontrol of a soil-borne pest for plants is a management option that is based on the identification and ecology of a naturally occurring soil predator or parasite that reduces the pest or

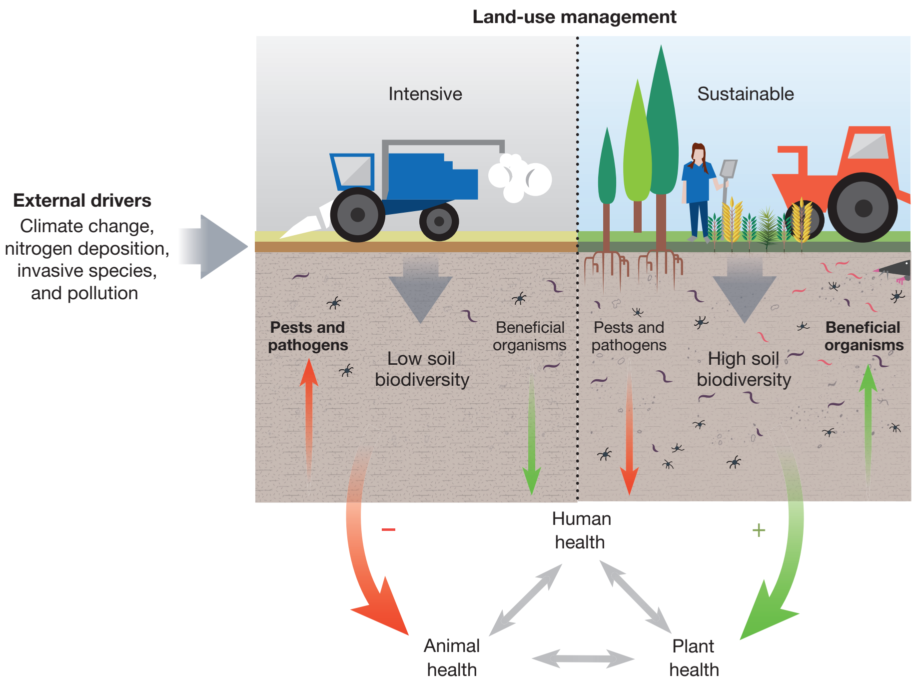


Figure 2 | A conceptual framework illustrating how decisions on land use and management are linked to human health through the effect on soil biodiversity. Soil biodiversity is strongly influenced by external drivers such as climate change and nitrogen deposition but also by land-use management. Land use such as agricultural intensification (left) can reduce the diversity and densities of beneficial organisms that control pests and pathogens, thereby negatively affecting the health of plants, animals and humans. Adopting less-intensive management practices (right) that enhance soil biodiversity can promote plant, animal and human health because the number of beneficial species will outweigh pests and pathogens. Moreover, soil biodiversity may help mitigate the impacts of external drivers of ecosystem functioning.

plant pathogen population and thereby enhances food quantity and nutrient content.

For example, the root weevil *Diaprepes abbreviatus* causes substantial damage to citrus plants. The damage is naturally controlled in some Florida soil habitats by species of indigenous soil entomopathogenic nematodes (EPN) that parasitize and kill the root weevil. A nematode species not native to Florida, *Steinernema riobrave*, is commercially available to control the root weevil in habitats where indigenous EPNs are less dominant, rather than using less effective, more costly chemical controls that move into groundwater and are harmful to human health[38,39]. In addition to augmenting soils with EPNs, managing soil in ways that promote native EPN prevalence and diversity are used. For example, maintaining good soil drainage, low pH, and adding sand to soils that are low in EPNs before planting young trees are practices used to conserve native EPN species in citrus groves. There are many other examples of biocontrol delivered through a healthy soil foodweb, such as the use of the bacterium *Pasteuria penetrans*, a pathogen of plant parasitic nematodes, and *Arthrobotrys anchonia,* a nematode-trapping fungus that kills plant parasitic nematodes[40–43].

As recent evidence suggests, it is not only single organisms that should be considered as valuable for controlling soil-borne pests and pathogens of humans, animals or plants. The immense diversity and abundance of organisms found belowground in concert contribute to the control of pests and pathogens (Box 1)[44–46]. Hence, control of soil-borne pathogens should not focus solely on specific beneficial soil predators or parasites, but rather on how a general increase in the complexity of soil biodiversity can reduce plant, animal and human diseases caused by soil-borne pathogens: the disease is suppressed as a result of the whole soil foodweb[47,48].

Although the direct link between biodiversity and disease suppression has not been well established in soil owing to the complex interactions that occur belowground, there is growing evidence for the aboveground world: disease risk in wildlife plants and humans rises with biodiversity loss[49–53]. For example, Johnson *et al.*[50] recently showed in wetland ecosystems that amphibian species richness moderated pathogen transfer and thereby limited disease prevalence in the animals. Soil biodiversity may similarly moderate the impacts of pests and pathogens both above- and belowground. One recent study[54] showed how increased microbial diversity reduced the success of a bacterial pathogen *in vitro*. Furthermore, temporary inhabitants of soil can also have positive functions in ecosystems and thereby indirectly benefit human health; for example, some bumblebees well known for their benefits as plant pollinators live temporarily in soils. Burrowing vertebrates such as voles and prairie dogs can also indirectly benefit plant, animal and human health by mixing and enriching organic matter and nutrients in soils[55,56]. These examples emphasize that many soil organisms contribute indirectly to soil functions that can ultimately benefit human health.

Maintaining soil biodiversity for health

To maintain soil biodiversity, it is essential to take into account the spatial distribution of belowground organisms. In recent years, the available information on biogeography of soil biodiversity has accelerated. Global distributions of soil taxa from microbes to larger animals shows that few species occur in all soils; instead many species are rare and show restricted distributions, often limited to particular soil types or geographical regions[57,58]. For example, some enchytraeid species occur primarily in rich Arctic peat soils, and many nematode and mite species are endemic to the Antarctic continent[58]. Likewise, a study of the soils of Central Park in New York City found almost as many distinct microbial communities and undescribed soil biodiversity (bacteria, archaea and eukarya) as occur in other global biomes[59]. Soil biodiversity, like soils themselves, is highly variable across fields and regions, highlighting the need to understand how soil communities organized in complex soil foodwebs differ spatially across a region and globally[57,58,60,61].

In the next sections, we outline how soil biodiversity influences the production of food, fibre and biomass, and the provision of clean water and air, and illustrate how improved management of soil biodiversity can reverse to some degree the negative impact of humans on the depletion of global resources (Fig. 2).

Food, fibre and biomass production

With the exception of hydroponic horticulture, all terrestrial crop production is soil-based[21]. Given that crops support most of the human population, sustainable use of our soils is essential for long-term human health. In agricultural systems, soil-borne pathogens can disrupt the metabolic flow of nutrients within plants, reduce plant above- and belowground biomass, including fruits and other edible plant parts, or even kill the plant entirely, all leading to the production of less nutritious food. In humid and cooler climates, earthworms have been shown to increase crop productivity[62], whereas termites can increase yields in warmer and

drier climates[63]. Bender and van der Heijden[64] showed that enriched soil life increased nutrient-use efficiency, plant nutrient uptake and thereby crop yields. Moreover, enhancing the soil foodweb structure influences the resistance and resilience of other terrestrial ecosystems[65,66], and this knowledge can be used to promote sustainable use of our soils[45,67].

Soil symbionts have a particularly important role in sustainable production. Symbiotic soil microbes are essential for nutrient supply[68] and can contribute to biofortification of plants for important micronutrients such as zinc[69]. Plant breeding and micronutrient fertilization are promising ways to address micronutrient deficiencies, but the bioavailability of the micronutrients is ultimately determined by soil microbial cycling of these micronutrients[16]. There is also evidence that fungi, in particular endophytes, promote plant stress tolerance[70].

It is evident from the above that soil biodiversity can play a crucial part in providing a more stable supply of food and a higher nutritional value of the food produced. However, the intensification of agricultural practices in the last century has ignored this role of soil biodiversity. The cornerstones of agricultural intensification—ploughing, and the application of agrochemicals and fertilizer—have been linked to a reduction of soil biodiversity[3,6]. We stress that these are beneficial practices that should not be abolished, but instead should be used at the right time, rate and place.

Air quality

Land-use change has been tied to the frequency of dust storms, emissions of greenhouse gases, and the release of volatile organic compounds and biota in air[7]. Soil bacteria, fungi and some invertebrates, such as nematodes and mites, are transported several hundreds to thousands of kilometres by wind[71,72]. The misuse of land—such as overly intensive ploughing, leaving extended areas bare and fallow, and burning plant biomass from fields—increases dust and the formation of particulate matter of less than 10 μm in size (PM10), with major consequences for human health in the form of respiratory problems, lung tissue damage, and even lung cancer[7,10]. Because soils are frequently polluted with heavy metals, harbour antibiotic-resistant organisms from animal feedlots, and contain pathogens for plants, animals and humans, the resulting dust can cause negative effects on human health[73,74].

An example is valley fever in the southwestern region of the USA; outbreaks are caused by a soil fungus, *Coccidioides immitis*, that normally decays organic matter and helps to stabilize the soil surface, thus minimizing soil erosion. However, when the soil is disturbed, such as by agricultural practices, the fungus produces windblown spores that can cause lung disease in animals and humans and at worst result in death[75,76]. In 2004, there were 6,000 cases of valley fever in the USA[74]. Surveillance of dust storms with land–atmosphere modelling and remote sensing of dust storms is under way to enhance the epidemiology and decrease the number of cases of valley fever[74].

Here again the link between agricultural intensification, soil biodiversity, and human health is clear; intensive agriculture disturbs the soil and negatively affects soil organisms, such as arbuscular mycorrhizae, saprotrophic fungi, and earthworms, that play a key part in stabilizing soil and thereby reduce the potential of dust formation[77,78]. Management options such as reduced tillage have been shown to reduce PM10 formation[73] and thus limit the risk of lung disease, cardiac arrhythmia, heart attacks and premature death[79]. Other ways to reduce dust and conserve soil stability and biodiversity include agroecological management practices, such as planting windbreaks, adding manure, incorporating cover crops and retaining crop residues[80].

Water quality

The provision of clean drinking water is increasingly compromised by pollution (such as from mining, landfills and agrochemicals)[81] and poor sanitation (contaminating drinking water with faecal-associated organisms)[16,82]. Moreover, land-use changes, especially those accompanying urbanization, affect the relationship between runoff versus infiltration of water with potential impacts on local surface water bodies, groundwater levels, areas downstream of point source pollution and the recharge of aquifers.

Soil biodiversity acts to enhance the structure of soils and thereby infiltration and percolation of water through the soil profile to (1) improve

BOX 2

The UN Sustainable Development Goals

The successful implementation of the UN Sustainable Development Goals[90] in September 2015 that are aimed at ending poverty and improving the lives of the poor, are tightly connected to maintaining the biodiversity of soils. Yet only four of the seventeen targets specifically mention soil: Goal 2 (to end hunger, achieve food security and improved nutrition, and promote sustainable agriculture (Target 2.4), Goal 3 (to ensure healthy lives and promote well-being for all at all ages (Target 3.9), Goal 12 (to ensure sustainable consumption and production patterns (Target 12.4), and Goal 15 (to protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification and halt and reverse land degradation, and halt biodiversity loss (Target 15.3).

For example, Goal 3 (Target 3.9) focuses on substantially reducing hazardous chemicals and air, water and soil pollution. However, the connection between managing land for enhanced soil biodiversity and meeting Sustainable Development Goals such as ending epidemics of tropical and other communicable diseases (Goal 3, Target 3.3) or sustainable management of water and sanitation (Goal 6) is not recognized or incorporated.

To achieve the Sustainable Development Goals we stress that it is not enough to aim towards improvement of a single benefit related to 'food' or 'air' or 'water' or 'disease' control, because all are simultaneously dependent on soils and soil biodiversity. We propose a multiple-benefit focus for sustaining soils, biodiversity and global health that addresses many of the Sustainable Development Goals—such as Goals 1, 2, 3, 6, 8, 11 13, 14 and 15—through the following means:

- Include soil biodiversity and human, plant and animal health experts in integrated collaborative research, management and policy efforts to sustainably manage soils, food, water and air for improving human health
- Develop a global database of soil biodiversity to facilitate integrated and predictive use by scientists and health experts
- Establish a global archive of samples for the future benefits of health researchers that captures the interactions of total soil biodiversity (bacteria, archaea, eukaryotes)
- Utilize existing, new and local knowledge on successful management of lands to promote new options for long-term maintenance and conservation of soil biodiversity and improving human health
- Include soil biodiversity as a criterion for determining wilderness and protected areas
- Focus research on conservation of soil biodiversity as a management tool to improve human health in the long-term
- Coordinate scientific societies and other global efforts to educate and communicate results to land and water managers, public and policy makers, such as through the Global Soil Biodiversity Initiative (https://globalsoilbiodiversity.org), global conventions and scientific societies
- Broaden the disciplines of human health and soil biodiversity linkages to include the combined expertise needed to address the multifaceted climate and global environmental changes and to meet the Sustainable Development Goals

water-use efficiency by crops, (2) limit the amount of agricultural runoff and associated contamination into adjacent land areas, and (3) filter out pathogens and contaminants by size exclusion, die-off and adsorption. Soil organisms can also degrade harmful pollutants and reduce the impact of poor sanitation[83,84]. For example, *Enterobacter cloacae*, an enteric bacterium found in soils and water, is an effective means of bioremediating selenium-contaminated agricultural drainage water[82]. Selenium, an essential micronutrient for humans, occurs in groundwater and can accumulate in irrigated river basins and evaporative ponds. Implementing additional measures such as reduced irrigation, sealing earthen irrigation canals, and rotational land fallowing can further enhance the management of excess selenium[85].

Outlook

It is clear that soil biodiversity represents an underutilized resource for sustaining or improving human health through better soil management. As indicated above, some agroecological management options are known to maintain and increase soil biodiversity for human, animal and plant health. However, further development of viable practices and especially the promotion of their use as broadly as possible is urgently needed.

How to best manage the world's lands for improved human health? Some basic guidelines for management of soil biodiversity are offered here. We suggest that a new approach for land use and management is required that acknowledges that soil biota act in concert to provide multiple benefits, even if these benefits are not easily observed. Moreover, increased soil foodweb complexity promotes resistance and resilience to perturbation and may buffer the impacts of extreme events.

Agroecological practices that enhance soil organic matter content and soil biodiversity can promote nutrient supply, water infiltration and well-structured soil. Effective management options for cropping systems include reduced tillage with residue retention and rotation, cover crop inclusion, integrated pest management, and integrated soil fertility management (such as the combination of chemical and organic fertilizer). Expanding plant species diversity in crop and/or land rotations and adding organic amendments to pastures can increase soil biodiversity and mimic better the natural soil foodweb[65,66,86]. Additionally, maintenance of soil biodiversity at the landscape level can be enhanced through buffer strips and riparian zones and land rotations. Drainage water management can reduce the movement of pollutants, agrochemicals and other contaminants to nearby landscapes[13]. Likewise, several forestry practices exist that promote soil biodiversity: re-established mixed deciduous forest stands in Europe were shown to have higher soil biodiversity than pure coniferous stands[87].

Management for conservation of land should include soil biodiversity as an important criterion in determining protected and wilderness areas, particularly in rapidly changing ecosystems, such as tropical forests, permafrost soils and alpine grasslands. Conservation of soil biodiversity should, in general terms, be based on existing knowledge of soil properties, the abundance, sizes and types of soil organisms, and vegetation. Nevertheless, conserving soil biodiversity could also be done through laboratory isolation of individual organisms or whole communities to maintain a reservoir of genetic and functional diversity appropriate for future disease prevention, biological technologies, and pharmaceuticals[88].

Soil archives that conserve live collections of interacting species of soil microbes and invertebrates in soil samples from different biomes are irreplaceable and essential; yet at present there are few such archives[88]. Given the growing global demands placed on limited productive land and the projected increases in infectious diseases, there is an urgent need to implement these and other conservation measures as a stockpile for the future.

Ideally, the practices and conservation strategies outlined above that enhance soil biodiversity for the maintenance of human health should be incorporated directly into land-, air- and water-use policies at global and regional levels and integrated with public health organizations such as the United Nations (UN) World Health Organization. Global conventions such as the UN Framework Convention on Climate Change, the UN Convention on Biological Diversity (CBD) and the UN Convention to Combat Desertification are all central to soils and global land use but often neglect soil biodiversity and our dependence on soil for human health, with the exception of the CBD[14] through the Food and Agricultural Organization (FAO). Through the Global Soil Partnership, the UN FAO brings together global institutions and other interested parties to coordinate agreements and international challenges related to soil sustainability. The Global Soil Partnership is advised on global soil issues by a scientific Intergovernmental Technical Panel on Soils. Likewise, progress towards the UN Sustainable Development Goals can be achieved by incorporating knowledge of soil biodiversity into a broader spectrum of benefits that improve human health (see Box 2; ref. 89). Importantly, the Global Soil Biodiversity Initiative was established as an independent scientific effort to provide information on soil biodiversity to policymakers and is preparing to publish the first Global Soil Biodiversity Atlas in collaboration with the European Union Joint Research Centre. The Global Soil Biodiversity Initiative (https://globalsoilbiodiversity.org) is also working to have soil biodiversity considered in current international initiatives such as the Intergovernmental Platform on Biodiversity and Ecosystem Services and Future Earth.

Fortunately, there is increased recognition that developing effective management tools for soil biodiversity requires active information transfer between scientists and policymakers with new policies formed on current evidence-based knowledge and local cultural knowledge[3,4]. However, we need to identify implementation mechanisms to encourage easier updates on best management practices and related policies to ensure long-term sustainable use of global lands under a changing global environment. This is particularly crucial given the rapid accumulation of new insights on how soil biodiversity can be managed to promote human health.

We are losing soils and soil biodiversity at a rapid pace, with substantial negative ramifications on human health worldwide. It is time to recognize and manage soil biodiversity as an underutilized resource for achieving long-term sustainability goals related to global human health, not only for improving soils, food security, disease control, water and air quality, but because biodiversity in soils is connected to all life and provides a broader, fundamental ecological foundation for working with other disciplines to improve human health.

Received 14 April; accepted 28 September 2015.
Published online 23 November 2015.

1. Bardgett, R. D. & van der Putten, W. H. Belowground biodiversity and ecosystem functioning. *Nature* **515,** 505–511 (2014).
2. Fisher, F. S., Bultman, M. W., Johnson, S. M., Pappagianis, D. & Zaborsky, E. *Coccidioides* niches and habitat parameters in the southwestern United States: a matter of scale. *Ann. NY Acad. Sci*. **1111,** 47–72 (2007).
3. Wall, D. H. *et al. Soil Ecology and Ecosystem Services* (Oxford Univ. Press, 2012).
4. Wall, D. H. & Six, J. Give soils their due. *Science* **347,** 695 (2015).
5. Haddad, N. M. *et al.* Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci. Adv.* **1,** e1500052 (2015).
6. Tsiafouli, M. A. *et al.* Intensive agriculture reduces soil biodiversity across Europe. *Glob. Change Biol.* **21,** 973–985 (2015).
This study, encompassing four agricultural regions across Europe, showed that increasing land-use intensity reduced soil foodweb diversity, functional diversity and taxonomic diversity.
7. Garrison, V. H. *et al.* African and Asian dust: from desert soils to coral reefs. *Bioscience* **53,** 469–480 (2003).
8. Park, J. W. *et al.* Effects of ambient particulate matter on peak expiratory flow rates and respiratory symptoms of asthmatics during Asian dust periods in Korea. *Respirology* **10,** 470–476 (2005).
9. Quinton, J. N., Govers, G., Van Oost, K. & Bardgett, R. D. The impact of agricultural soil erosion on biogeochemical cycling. *Nature Geosci*. **3,** 311–314 (2010).
10. Schenker, M. Exposures and health effects from inorganic agricultural dusts. *Environ. Health Perspect.* **108,** 661–664 (2000).
11. Dominati, E., Patterson, M. & Mackay, A. A framework for classifying and quantifying the natural capital and ecosystem services of soils. *Ecol. Econ.* **69,** 1858–1868 (2010).
12. Thiele-Bruhn, S., Bloem, J., de Vries, F. T., Kalbitz, K. & Wagg, C. Linking soil biodiversity and agricultural soil management. *Curr. Opin. Environ. Sustainability* **4,** 523–528 (2012).

13. Nielsen, U. N., Wall, D. H. & Six, J. Soil biodiversity and the environment. *Annu. Rev. Environ. Resour.* **40,** 63–90 (2015).
14. World Health Organization and Secretariat of the Convention on Biological Diversity. *Connecting Global Priorities: Biodiversity and Human Health. A State of Knowledge Review* https://www.cbd.int/health/SOK-biodiversity-en.pdf (WHO, 2015).
15. Brevik, E. C. & Burgess, L. C. The 2012 fungal meningitis outbreak in the United States: connections between soils and human health. *Soil Horizons* **54,** 1–4 (2013).
16. Oliver, M. A. & Gregory, P. J. Soil, food security and human health: a review. *Eur. J. Soil Sci.* **66,** 257–276 (2015).
17. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517,** 455–459 (2015).
18. Ferris, H. & Tuomisto, H. Unearthing the role of biological diversity in soil health. *Soil Biol. Biochem.* **85,** 101–109 (2015).
19. Brevik, E. C. & Burgess, L. C. *Soils and Human Health* (CRC Press, 2012).
20. Bultman, M. W., Fisher, F. S. & Pappagianis, D. in *Essentials of Medical Geology* (ed. O. Selinus) Ch. 20 (Springer, 2013).
21. Pepper, I. L., Gerba, C. P., Newby, D. T. & Rice, C. W. Soil: a public health threat or savior? *Crit. Rev. Environ. Sci. Technol.* **39,** 416–432 (2009).
22. Brevik, E. C. & Sauer, T. J. The past, present, and future of soils and human health studies. *Soil* **1,** 35–46 (2015).
23. Myers, S. S. & Patz, J. A. Emerging threats to human health from global environmental change. *Annu. Rev. Environ. Resour.* **34,** 223–252 (2009).
24. Berg, G., Eberl, L. & Hartmann, A. The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. *Environ. Microbiol.* **7,** 1673–1685 (2005).
25. Ganz, H. H. *et al.* Interactions between *Bacillus anthracis* and plants may promote anthrax transmission. *PLoS Negl. Trop. Dis.* **8,** http://dx.doi.org/10.1371/journal.pntd.0002903 (2014).
26. Smith, K. L. *et al. Bacillus anthracis* diversity in Kruger National Park. *J. Clin. Microbiol.* **38,** 3780–3784 (2000).
27. Schär, F. *et al. Strongyloides stercoralis*: global distribution and risk factors. *PLoS Negl. Trop. Dis.* **7,** http://dx.doi.org/10.1371/journal.pntd.0002288 (2013).
28. Khieu, V. *et al.* High prevalence and spatial distribution of *Strongyloides stercoralis* in rural Cambodia. *PLoS Negl. Trop. Dis.* **8,** http://dx.doi.org/10.1371/journal.pntd.0002854 (2014).
29. de Silva, N. R. *et al.* Soil-transmitted helminth infections: updating the global picture. *Trends Parasitol.* **19,** 547–551 (2003).
30. Kay, A. B. Overview of "Allergy and allergic diseases: with a view to the future". *Br. Med. Bull.* **56,** 843–864 (2000).
31. Matricardi, P. M. & Bonini, S. High microbial turnover rate preventing atopy: a solution to inconsistencies impinging on the hygiene hypothesis? *Clin. Exp. Allergy* **30,** 1506–1510 (2000).
32. Rook, G. A. W. 99th Dahlem conference on infection, inflammation and chronic inflammatory disorders: Darwinian medicine and the 'hygiene' or 'old friends' hypothesis. *Clin. Exp. Immunol.* **160,** 70–79 (2010).
33. Hanski, I. *et al.* Environmental biodiversity, human microbiota, and allergy are interrelated. *Proc. Natl Acad. Sci. USA* **109,** 8334–8339 (2012).
This study provides evidence that people living near environmentally diverse areas had less propensity for allergies because of a greater diversity of commensal bacteria on their skin, most of which are also found in soil and vegetation.
34. Haahtela, T. *et al.* The Finnish Allergy Programme 2008–2018—scientific rationale and practical implementation. *Asia Pacific Allergy* **2,** 275–279 (2012).
35. Ruokolainen, L. *et al.* Green areas around homes reduce atopic sensitization in children. *Allergy* **70,** 195–202 (2015).
36. Prichard, R. in *Antimicrobial Drug Resistance* (ed. Mayers, D. L.) 621–628 (Springer, 2009).
37. Corbett, C. J. *et al.* The effectiveness of faecal removal methods of pasture management to control the cyathostomin burden of donkeys. *Parasites Vectors* **7,** 48 (2014).
38. Epstein, L. Fifty years since *Silent Spring*. *Annu. Rev. Phytopathol.* **52,** 377–402 (2014).
39. Campos-Herrera, R., El-Borai, F. E. & Duncan, L. W. in *Nematode Pathogenesis of Insects and Other Pests* (ed. Campos-Herrera, R.) (Springer, 2015).
40. Charles, L. *et al.* Phylogenetic analysis of *Pasteuria penetrans* by use of multiple genetic loci. *J. Bacteriol.* **187,** 5700–5708 (2005).
41. Gray, N. F. Ecology of nematophagous fungi: *Panagrellus redivivus* as the target organism. *Plant Soil* **73,** 293–297 (1983).
42. Stirling, G. R. in *Biological Control of Plant-Parasitic Nematodes: Building Coherence Between Microbial Ecology and Molecular Mechanisms* (eds Davies, K. G. & Spiegel, Y.) 1–38 (Springer, 2011).
43. Stirling, G. R. *Biological Control of Plant-Parasitic Nematodes: Soil Ecosystem Management Sustainable Agriculture* 2nd edn (CABI, 2014).
44. Costa, S. R., Kerry, B. R., Bardgett, R. D. & Davies, K. G. Interactions between nematodes and their microbial enemies in coastal sand dunes. *Oecologia* **170,** 1053–1066 (2012).
45. Ferris, H. *et al.* Diversity and complexity complement apparent competition: nematode assemblages in banana plantations. *Acta Oecol.* **40,** 11–18 (2012).
46. Sánchez-Moreno, S. & Ferris, H. Suppressive service of the soil food web: effects of environmental management. *Agric. Ecosyst. Environ.* **119,** 75–87 (2007).
This study showed that the prevalence of predator and omnivorous nematodes, which suppressed plant parasitic nematodes, was higher in soils with more complex foodwebs.
47. Penton, C. R. *et al.* Fungal community structure in disease suppressive soils assessed by 28S LSU gene sequencing. *PLoS ONE* **9,** http://dx.doi.org/10.1371/journal.pone.0093893 (2014).
48. Weller, D. M., Raaijmakers, J. M., Gardener, B. B. M. & Thomashow, L. S. Microbial populations responsible for specific soil suppressiveness to plant pathogens. *Annu. Rev. Phytopathol.* **40,** 309–348 (2002).
49. Johnson, P. T. J. *et al.* Species diversity reduces parasite infection through cross-generational effects on host abundance. *Ecology* **93,** 56–64 (2012).
50. Johnson, P. T. J., Preston, D. L., Hoverman, J. T. & Richgels, K. L. D. Biodiversity decreases disease through predictable changes in host community competence. *Nature* **494,** 230–233 (2013).
This laboratory and field study showed that a richer host diversity reduced transmission of a parasite (the trematode *Ribeiroia ondatrae*) and reduced amphibian disease.
51. Keesing, F., Holt, R. D. & Ostfeld, R. S. Effects of species diversity on disease risk. *Ecol. Lett.* **9,** 485–498 (2006).
52. Searle, C. L., Biga, L. M., Spatafora, J. W. & Blaustein, A. R. A dilution effect in the emerging amphibian pathogen *Batrachochytrium dendrobatidis*. *Proc. Natl Acad. Sci. USA* **108,** 16322–16326 (2011).
53. Suzán, G. *et al.* Experimental evidence for reduced rodent diversity causing increased hantavirus prevalence. *PLoS ONE* **4,** http://dx.doi.org/10.1371/journal.pone.0005461 (2009).
54. van Elsas, J. D. *et al.* Microbial diversity determines the invasion of soil by a bacterial pathogen. *Proc. Natl Acad. Sci. USA* **109,** 1159–1164 (2012).
Experimental results showed that in soils with greater microbial diversity invading bacteria have a lower survival rate.
55. Andersen, D. C. Belowground herbivory in natural communities—a review emphasizing fossorial animals. *Q. Rev. Biol.* **62,** 261–286 (1987).
56. Gehring, C. A., Wolf, J. E. & Theimer, T. C. Terrestrial vertebrates promote arbuscular mycorrhizal fungal diversity and inoculum potential in a rain forest soil. *Ecol. Lett.* **5,** 540–548 (2002).
57. Bates, S. T. *et al.* Global biogeography of highly diverse protistan communities in soil. *ISME J.* **7,** 652–659 (2013).
58. Wu, T., Ayres, E., Bardgett, R. D., Wall, D. H. & Garey, J. R. Molecular study of worldwide distribution and diversity of soil animals. *Proc. Natl Acad. Sci. USA* **108,** 17720–17725 (2011).
59. Ramirez, K. S. *et al.* Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proc. R. Soc. B* **281,** http://dx.doi.org/10.1098/rspb.2014.1988 (2014).
60. Lauber, C. L., Ramirez, K. S., Aanderud, Z., Lennon, J. & Fierer, N. Temporal variability in soil microbial communities across land-use types. *ISME J.* **7,** 1641–1650 (2013).
61. Tedersoo, L. *et al.* Global diversity and geography of soil fungi. *Science* **346,** http://dx.doi.org/10.1126/science.1256688 (2014).
This analysis indicates that the distribution and species richness of fungi is mostly determined by climate and not related to plant diversity on a global scale, except for root-related ectomycorrizal fungi; biogeographical comparisons of continents indicate efficient dispersal mechanisms for fungi compared to larger organisms.
62. van Groenigen, J. W. *et al.* Earthworms increase plant production: a meta-analysis. *Sci. Rep.* **4,** http://dx.doi.org/10.1038/srep06365 (2014).
63. Evans, T. A., Dawes, T. Z., Ward, P. R. & Lo, N. Ants and termites increase crop yield in a dry climate. *Nature Commun.* **2,** http://dx.doi.org/10.1038/ncomms1257 (2011).
64. Bender, S. F. & van der Heijden, M. G. A. Soil biota enhance agricultural sustainability by improving crop yield, nutrient uptake and reducing nitrogen leaching losses. *J. Appl. Ecol.* **52,** 228–239 (2015).
65. de Vries, F. T. *et al.* Soil food web properties explain ecosystem services across European land use systems. *Proc. Natl Acad. Sci. USA* **110,** 14296–14301 (2013).
66. de Vries, F. T. *et al.* Land use alters the resistance and resilience of soil food webs to drought. *Nature Clim. Change* **2,** 276–280 (2012).
This study compared nitrogen retention in extensive versus intensive grasslands and showed that species-rich extensively managed grasslands had greater soil nitrogen retention due to a higher fungal:bacterial abundance ratio compared to intensively managed grasslands.
67. Wall, D. H., Bardgett, R. D. & Kelly, E. F. Biodiversity in the dark. *Nature Geosci.* **3,** 297–298 (2010).
68. Rillig, M. C. & Mummey, D. L. Mycorrhizas and soil structure. *New Phytol.* **171,** 41–53 (2006).
69. Schulin, R., Khoshgoftarmanesh, A., Afyuni, M., Nowack, B. & Frossard, E. in *Development and Uses of Biofortified Agricultural Products* (eds Banuelos, G. S. & Lin, Z.-Q.) Ch. 6 (CRC Press, 2008).
70. Rodriguez, R. J. *et al.* Stress tolerance in plants via habitat-adapted symbiosis. *ISME J.* **2,** 404–416 (2008).
71. Brown, J. K. M. & Hovmøller, M. S. Epidemiology—aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* **297,** 537–541 (2002).
72. Nkem, J. N. *et al.* Wind dispersal of soil invertebrates in the McMurdo Dry Valleys, Antarctica. *Polar Biol.* **29,** 346–352 (2006).

73. Madden, N. M., Southard, R. J. & Mitchell, J. P. Conservation tillage reduces PM10 emissions in dairy forage rotations. *Atmos. Environ.* **42,** 3795–3808 (2008).
74. Sprigg, W. A. *et al.* Regional dust storm modeling for health services: the case of valley fever. *Aeolian Res.* **14,** 53–73 (2014).
75. Nguyen, C. *et al.* Recent advances in our understanding of the environmental, epidemiological, immunological, and clinical dimensions of coccidioidomycosis. *Clin. Microbiol. Rev.* **26,** 505–525 (2013).
76. Tabor, J. A., O'Rourke, M. K., Lebowitz, M. D. & Harris, R. B. Landscape-epidemiological study design to investigate an environmentally based disease. *J. Expo. Sci. Environ. Epidemiol.* **21,** 197–211 (2011).
77. Frey, S. D., Elliott, E. T. & Paustian, K. Bacterial and fungal abundance and biomass in conventional and no-tillage agroecosystems along two climatic gradients. *Soil Biol. Biochem.* **31,** 573–585 (1999).
78. Six, J., Frey, S. D., Thiet, R. K. & Batten, K. M. Bacterial and fungal contributions to carbon sequestration in agroecosystems. *Soil Sci. Soc. Am. J.* **70,** 555–569 (2006).
79. Anderson, J. O., Thundiyil, J. G. & Stolbach, A. Clearing the air: a review of the effects of particulate matter air pollution on human health. *J. Med. Toxicol.* **8,** 166–175 (2012).
This review provides an analysis of the complexity of particulate matter air pollution and its effects on human health.
80. Nordstrom, K. F. & Hotta, S. Wind erosion from cropland solutions in the USA: a review of problems, and prospects. *Geoderma* **121,** 157–167 (2004).
81. Alavanja, M. C. R., Ross, M. K. & Bonner, M. R. Increased cancer burden among pesticide applicators and others due to pesticide exposure. *CA Cancer J. Clin.* **63,** 120–142 (2013).
82. Frankenberger, W. T. & Arshad, M. Bioremediation of selenium-contaminated sediments and water. *Biofactors* **14,** 241–254 (2001).
83. Abraham, J. & Silambarasan, S. Biodegradation of chlorpyrifos and its hydrolyzing metabolite 3,5,6-trichloro-2-pyridinol by *Sphingobacterium* sp JAS3. *Process Biochem.* **48,** 1559–1564 (2013).
84. Rayu, S., Karpouzas, D. G. & Singh, B. K. Emerging technologies in bioremediation: constraints and opportunities. *Biodegradation* **23,** 917–926 (2012).
85. Bailey, R. T., Romero, E. C. & Gates, T. K. Assessing best management practices for remediation of selenium loading in groundwater to streams in an irrigated region. *J. Hydrol.* **521,** 341–359 (2015).
86. Tiemann, L., Grandy, A., Atkinson, E., Marin-Spiotta, E. & McDaniel, M. Crop rotational diversity enhances belowground communities and functions in an agrosystem. *Ecol. Lett.* **18,** 761–771 (2015).
87. Chauvat, M., Titsch, D., Zaytsev, A. S. & Wolters, V. Changes in soil faunal assemblages during conversion from pure to mixed forest stands. *For. Ecol. Manage.* **262,** 317–324 (2011).
88. Cary, S. C. & Fierer, N. The importance of sample archiving in microbial ecology. *Nature Rev. Microbiol.* **12,** 789–790 (2014).
89. Jeffery, S. & van der Putten, W. H. *Soil-Borne Human Diseases* 1–56, http://dx.doi.org/10.2788/37199 (Joint Research Centre Scientific and Technical Reports, European Commission, 2011).
90. United Nations Sustainable Development Goals. Open Working Group Proposal for Sustainable Development Goals, https://sustainabledevelopment.un.org/focussdgs.html (UN, 2014).

Acknowledgements We appreciate the encouragement of J. Lehmann to write this article, and acknowledge the comments of B. Adams, L. Duncan, A. Franco, T. Fraser, M. Knox, K. Pintauro, A. Shaw, A. Weller and S. Vandewoude. D.H.W. acknowledges the Winslow Foundation.

Author Contributions D.H.W., J.S. and U.N.N. contributed equally to the planning and writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.H.W. (diana.wall@colostate.edu).

doi:10.1038/nature15750

Measuring entanglement entropy in a quantum many-body system

Rajibul Islam[1], Ruichao Ma[1], Philipp M. Preiss[1], M. Eric Tai[1], Alexander Lukin[1], Matthew Rispoli[1] & Markus Greiner[1]

Entanglement is one of the most intriguing features of quantum mechanics. It describes non-local correlations between quantum objects, and is at the heart of quantum information sciences. Entanglement is now being studied in diverse fields ranging from condensed matter to quantum gravity. However, measuring entanglement remains a challenge. This is especially so in systems of interacting delocalized particles, for which a direct experimental measurement of spatial entanglement has been elusive. Here, we measure entanglement in such a system of itinerant particles using quantum interference of many-body twins. Making use of our single-site-resolved control of ultracold bosonic atoms in optical lattices, we prepare two identical copies of a many-body state and interfere them. This enables us to directly measure quantum purity, Rényi entanglement entropy, and mutual information. These experiments pave the way for using entanglement to characterize quantum phases and dynamics of strongly correlated many-body systems.

Entangled quantum objects[1] are correlated in ways that reject the principle of local realism. In few-level quantum systems, entangled states have been investigated extensively as a means of studying the foundations of quantum mechanics[2] and as a resource for quantum information applications[3]. Recently, it was realized that the concept of entanglement has broad impact in many areas of quantum many-body physics, ranging from condensed matter[4] to high-energy field theory[5] and quantum gravity[6]. In this general context, entanglement is most often quantified by the entropy of entanglement[1] that arises in a subsystem when the information about the remaining system is ignored. This entanglement entropy exhibits qualitatively different behaviour from that of classical entropy and has been used in theoretical physics to probe various properties of many-body systems. In condensed matter physics, for example, the scaling behaviour[7] of entanglement entropy allows phases to be distinguished that cannot be characterized by symmetry properties, such as topological states of matter[8–10] and spin liquids[11,12]. Entanglement entropy can be used to probe quantum criticality[13] and non-equilibrium dynamics[14,15], and to determine whether efficient numerical techniques for computing many-body physics exist[16].

Despite the growing importance of entanglement in theoretical physics, current condensed matter experiments do not have a direct probe with which to detect and measure entanglement. Synthetic quantum systems such as cold atoms[17,18], photonic networks[19], and some microscopic solid state devices[20] have unique advantages: in such systems control and detection of single particles are possible, they provide experimental access to relevant dynamical timescales, and they are isolated from the environment. In these systems, specific entangled states of few qubits, such as the highly entangled Greenberger–Horne–Zeilinger (GHZ) state[21] have been experimentally created and detected using witness operators[22]. However, entanglement witnesses are state specific. For arbitrary states, an exhaustive method of reconstructing the entire quantum state by tomography[23] can be used to measure entanglement. This has been accomplished in small systems of photonic qubits[24] and trapped ion spins[25], but there is no known way to perform tomography for systems involving itinerant delocalized particles. With multiple copies of a system, however, one can use quantum many-body interference to quantify entanglement even in itinerant systems[15,26,27].

In this work, we take advantage of the precise control and readout afforded by our quantum gas microscope[28] to prepare and interfere two identical copies of a four-site Bose–Hubbard system. This many-body quantum interference enables us to measure quantities that are not directly accessible in a single system (without tomography), for example, quadratic functions of the density matrix[15,26,27,29–32]. Such nonlinear functions can reveal entanglement[1]. In our system, we directly measure the quantum purity, Rényi entanglement entropy, and mutual information to probe the entanglement in site occupation numbers.

Bipartite entanglement

To detect entanglement in our system, we use a fundamental property of entanglement between two subsystems (bipartite entanglement): ignoring information about one subsystem results in the other becoming a classical mixture of pure quantum states. This classical mixture in a density matrix ρ can be quantified by measuring the quantum purity, defined as $\mathrm{Tr}(\rho^2)$. For a pure quantum state the density matrix is a projector and $\mathrm{Tr}(\rho^2)=1$, whereas for a mixed state $\mathrm{Tr}(\rho^2)<1$. In the case of a product state, the subsystems A and B of a many-body system AB described by a separable wavefunction $|\psi_{\mathrm{AB}}\rangle$ (Fig. 1) are individually pure as well, that is, $\mathrm{Tr}(\rho_{\mathrm{A}}^2)=\mathrm{Tr}(\rho_{\mathrm{B}}^2)=\mathrm{Tr}(\rho_{\mathrm{AB}}^2)=1$. Here the reduced density matrix of A is $\rho_{\mathrm{A}}=\mathrm{Tr}_{\mathrm{B}}(\rho_{\mathrm{AB}})$, where $\rho_{\mathrm{AB}}=|\psi_{\mathrm{AB}}\rangle\langle\psi_{\mathrm{AB}}|$ is the density matrix of the full system. Tr_{B} indicates tracing over or ignoring all information about the subsystem B. For an entangled state, the subsystems become less pure compared to the full system as the correlations between A and B are ignored in the reduced density matrix, $\mathrm{Tr}(\rho_{\mathrm{A}}^2)=\mathrm{Tr}(\rho_{\mathrm{B}}^2)<\mathrm{Tr}(\rho_{\mathrm{AB}}^2)=1$. Even if the many-body state is mixed ($\mathrm{Tr}(\rho_{\mathrm{AB}}^2)<1$), it is still possible to measure entanglement between the subsystems[1]. It is sufficient[33] to prove this entanglement by showing that the subsystems are less pure than the full system, that is:

$$\begin{aligned}\mathrm{Tr}(\rho_{\mathrm{A}}^2)&<\mathrm{Tr}(\rho_{\mathrm{AB}}^2)\\ \mathrm{Tr}(\rho_{\mathrm{B}}^2)&<\mathrm{Tr}(\rho_{\mathrm{AB}}^2)\end{aligned} \tag{1}$$

These inequalities provide a powerful tool with which to detect entanglement in the presence of experimental imperfections. Furthermore, quantitative bounds on the entanglement present in a mixed many-body state can be obtained from these state purities[34].

[1]Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA.

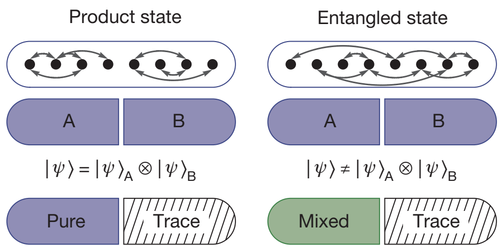


Figure 1 | Bipartite entanglement and partial measurements. A generic pure quantum many-body state has quantum correlations (shown as arrows) between different parts. If the system is divided into two subsystems A and B, the subsystems will be bipartite entangled with each other when there are quantum correlations between them (right column). Only when there is no bipartite entanglement present, the partitioned system $|\psi_{AB}\rangle$ can be described as a product of subsystem states $|\psi_A\rangle$ and $|\psi_B\rangle$ (left column). A path for measuring the bipartite entanglement emerges from the concept of partial measurements: ignoring all information about subsystem B (indicated as 'Trace') will put subsystem A into a statistical mixture, to a degree given by the amount of bipartite entanglement present. Finding ways of measuring the many-body quantum state purity of the system and comparing that of its subsystems would then enable measurements of entanglement. For an entangled state, the subsystems will have less purity than the full system.

Equation (1) can be framed in terms of entropic quantities[1,33]. A particularly useful and well studied quantity is the nth-order Rényi entropy:

$$S_n(\mathrm{A}) = \frac{1}{1-n} \log \mathrm{Tr}(\rho_\mathrm{A}^n) \quad (2)$$

From equation (2), we see that the second-order ($n=2$) Rényi entropy and purity are related by $S_2(\mathrm{A}) = -\log \mathrm{Tr}(\rho_\mathrm{A}^2)$. $S_2(\mathrm{A})$ provides a lower bound[15] for the von Neumann entanglement entropy $S_{\mathrm{VN}}(\mathrm{A}) = S_1(\mathrm{A}) = -\mathrm{Tr}(\rho_\mathrm{A} \log \rho_\mathrm{A})$, which has been extensively studied theoretically. The Rényi entropies are rapidly gaining importance in theoretical condensed matter physics because they can be used to extract information about the "entanglement spectrum"[35], thus providing more complete knowledge about the quantum state than just the von Neuman entropy. In terms of the second-order Rényi entropy, the conditions sufficient to demonstrate entanglement[1,33] become $S_2(\mathrm{A}) > S_2(\mathrm{AB})$, and $S_2(\mathrm{B}) > S_2(\mathrm{AB})$, that is, the subsystems have more entropy than the full system. These entropic inequalities are more powerful in detecting certain entangled states than other inequalities such as the Clauser–Horne–Shimony–Holt (CHSH) inequality[30,33].

Measurement of quantum purity

The quantum purity and hence the second-order Rényi entropy can be directly measured by interfering two identical and independent copies of the quantum state on a 50%–50% beam splitter[15,26,27,30]. For two identical copies of a bosonic Fock state, the output ports always have even particle numbers, as illustrated in Fig. 2a. This is due to the destructive interference of all odd outcomes. If the system is composed of multiple modes, such as internal spin states or various lattice sites the expectation value of the total number parity $P_i = \prod_k p_i^{(k)}$ is equal to unity in the output ports $i=1, 2$. Here the parity for mode k is $p_i^{(k)} = \pm 1$ for even or odd numbers of particles, respectively.

The well known Hong–Ou–Mandel (HOM) interference of two identical single photons[36] is a special case of this scenario. Here a pair of indistinguishable photons incident upon different input ports of a 50%–50% beam splitter interfere such that both photons always exit from the same output port. In general, the average parity measured in the many-body bosonic interference on a beam splitter probes the quantum state overlap (Supplementary Information) between the two copies, $\langle P_i \rangle = \mathrm{Tr}(\rho_1 \rho_2)$, where ρ_1 and ρ_2 are the density matrices of the two copies respectively and $\langle ... \rangle$ denotes averaging over repeated experimental realizations, as shown in Fig. 2b. Hence, for two identical

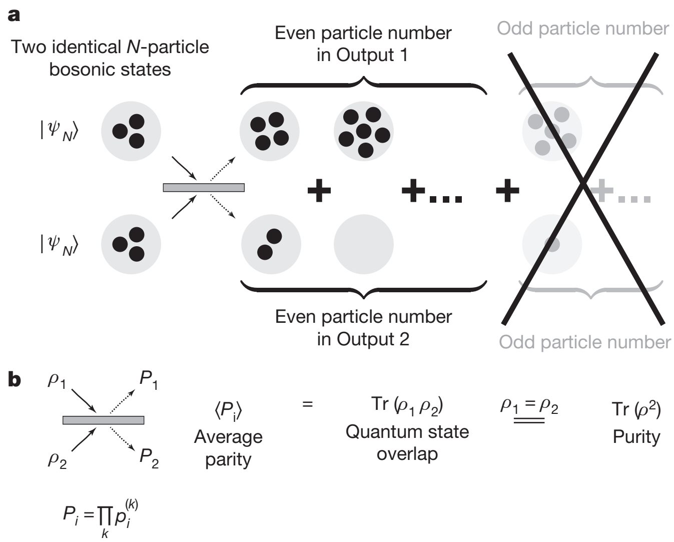


Figure 2 | Measurement of quantum purity with many-body bosonic interference of quantum twins. a, When two N-particle bosonic systems that are in identical pure quantum states are interfered on a 50%–50% beam splitter, they always produce output states with an even number of particles in each copy. This is due to the destructive interference of odd outcomes and represents a generalized HOM interference, in which two identical photons always appear in pairs after interfering on a beam splitter. **b**, If the input states ρ_1 and ρ_2 are not perfectly identical or not perfectly pure, the interference contrast is reduced. In this case the expectation value of the parity of particle number $\langle P_i \rangle$ in either output ($i=1, 2$) measures the quantum state overlap between the two input states. For two identical input states $\rho_1 = \rho_2$, the average parity $\langle P_i \rangle$ therefore directly measures the quantum purity of the states. We assume only that the input states have no relative macroscopic phase relationship.

systems, that is, for $\rho_1 = \rho_2 = \rho$, the average parity for both output ports ($i=1, 2$) equals the quantum purity of the many-body state[15,26,27]:

$$\langle P_i \rangle = \mathrm{Tr}(\rho^2) \quad (3)$$

Equation (3) represents the most important theoretical foundation behind this work—it connects a quantity depending on quantum coherences in the system to a simple observable in the number of particles. It holds even without fixed particle number, as long as there is no definite phase relationship between the copies (Supplementary Information). From equations (1) and (3), detecting entanglement in an experiment is thus reduced to simply measuring the average particle number parity in the output ports of the multi-mode beam splitter.

We probe entanglement formation in a system of interacting ^{87}Rb atoms on a one-dimensional optical lattice with a lattice constant of 680 nm. The dynamics of atoms in the lattice is described by the Bose–Hubbard Hamiltonian:

$$H = -J \sum_{\langle i,j \rangle} a_i^\dagger a_j + \frac{U}{2} \sum_i n_i(n_i - 1) \quad (4)$$

where $a_i^\dagger, a_i$ and $n_i = a_i^\dagger a_i$ are the bosonic creation, annihilation, and the number operators at site i, respectively. The atoms tunnel between neighbouring lattice sites (indicated by $\langle i, j \rangle$) with a rate J and experience an onsite repulsive interaction energy U. Planck's constant h is set to 1 and hence both J and U are expressed in hertz. The dimensionless parameter U/J is controlled by the depth of the optical lattice. Additionally, we can superimpose an arbitrary optical potential with the resolution of a single lattice site by using a spatial light modulator as an amplitude hologram through a high-resolution microscope (Supplementary Information). This microscope also allows us to image the number parity of each lattice site independently[28].

To initialize two independent and identical copies of a state with fixed particle number N, we start with a low-entropy two-dimensional Mott insulator with unity filling in the atomic limit[28] and deterministically retain a plaquette of $2 \times N$ atoms while removing all others (Supplementary Information). This is illustrated in Fig. 3a. The plaquette of $2 \times N$ atoms contains two copies (along the y direction) of an N-atom one-dimensional system (along the x direction), with $N = 4$ in this figure. The desired quantum state is prepared by manipulating the depth of the optical lattice along x, varying the parameter U/J_x, where J_x is the tunnelling rate along x. A box potential created by the spatial light modulator is superimposed onto this optical lattice to constrain the dynamics to the sites within each copy. During the state preparation, a deep lattice barrier separates the two copies and makes them independent of each other.

The beam splitter operation required for the many-body interference is realized in a double-well potential along y. The dynamics of atoms in the double well is likewise described by the Bose–Hubbard Hamiltonian, equation (4). A single atom, initially localized in one well, coherently oscillates between the wells with a Rabi frequency of $J = J_y$ (oscillation frequency in the amplitude). At discrete times during this evolution, $t = t_{\mathrm{BS}}^{(n)} = \frac{2n-1}{8J_y}$, with $n = 1, 2, \ldots$, the atom is delocalized equally over the two wells with a fixed phase relationship. Each of these times realizes a beam splitter operation, for which the same two wells serve as the input ports at time $t = 0$ and output ports at time $t = t_{\mathrm{BS}}^{(n)}$. Two indistinguishable atoms with negligible interaction strength ($U/J_y \ll 1$) in this double well will interfere as they tunnel. The dynamics of two atoms in the double well is demonstrated in Fig. 3b in terms of the joint probability $\mathcal{P}(1, 1)$ of finding them in separate wells versus the normalized time $J_y t$. The joint probability $\mathcal{P}(1, 1)$ oscillates at a frequency of 772(16) Hz $= 4J_y$, with a contrast of 95(3)%. At the beam splitter times, $t = t_{\mathrm{BS}}^{(n)}$, $\mathcal{P}(1, 1) \approx 0$. The first beam splitter time, $t_{\mathrm{BS}} \equiv t_{\mathrm{BS}}^{(1)} = \frac{1}{8J_y}$ is used for all the following experiments, with $\mathcal{P}(1, 1) = 0.05(2)$. This is a signature of bosonic interference of two indistinguishable particles[37,38], akin to the photonic HOM interference[36]. This high interference contrast indicates the near-perfect suppression of classical noise and fluctuations and includes an expected 0.6% reduction due to finite interaction strength ($U/J_y \approx 0.3$). The results from this interference can be interpreted as a measurement of the quantum purity of the initial Fock state as measured from the average parity (equation (3)), $\langle P_i \rangle = 1 - 2 \times \mathcal{P}(1, 1) = 0.90(4)$, where $i = 1, 2$ are the two copies.

Entanglement in the ground state

The Bose–Hubbard model provides an interesting system in which to investigate entanglement. In optical lattice systems, a lower bound of the spatial entanglement has been previously estimated from time-of-flight measurements[39] and entanglement dynamics in spin degrees of freedom has been investigated with partial state reconstruction[40]. Here, we directly measure entanglement in real space occupational particle number in a site-resolved way. In the strongly interacting atomic limit of $U/J_x \gg 1$, the ground state is a Mott insulator corresponding to a Fock state of one atom at each lattice site. The quantum state has no spatial entanglement with respect to any partitioning in this phase—it is in a product state of the Fock states. As the interaction strength is reduced adiabatically, atoms begin to tunnel across the lattice sites, and ultimately the Mott insulator melts into a superfluid with a fixed atom number. The delocalization of atoms creates entanglement between spatial subsystems. This entanglement originates[41,42] from correlated fluctuations in the number of particles between the subsystems due to the super-selection rule that the total particle number in the full system is fixed, as well as coherence between various configurations without any such fluctuation.

To probe the emergence of entanglement, we first prepare the ground state of equation (4) in both copies by adiabatically lowering the optical

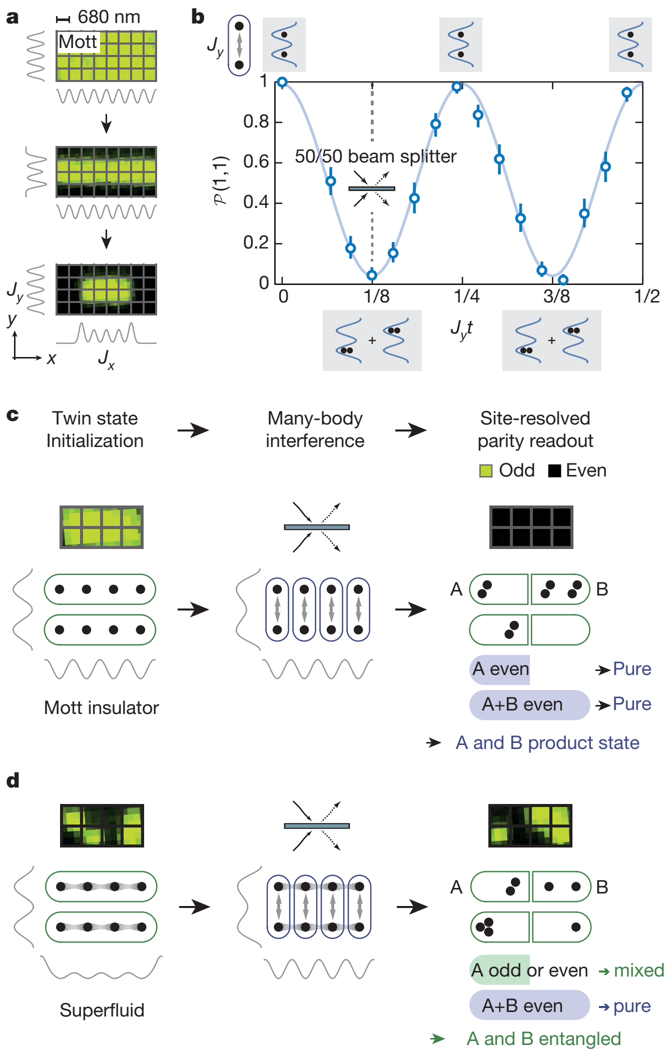


Figure 3 | Many-body interference to probe entanglement in optical lattices. a, A high-resolution microscope is used to directly image the number parity of ultracold bosonic atoms on each lattice site (in the raw images, green represents odd and black represents even). Two adjacent one-dimensional lattices are created by combining an optical lattice and potentials created by a spatial light modulator. We initialize two identical many-body states by filling the potentials from a low-entropy two-dimensional Mott insulator. The tunnelling rates J_x and J_y can be tuned independently by changing the depth of the potential. **b**, The atomic beam splitter operation is realized in a tunnel-coupled double-well potential. An atom, initially localized in one of the wells, delocalizes with equal probability into both the wells by this beam splitter. Here, we show the atomic analogue of the HOM interference of two states. The joint probability $\mathcal{P}(1, 1)$ measures the probability of coincidence detection of the atoms in separate wells as a function of normalized tunnel time $J_y t$, with the single particle tunnelling $J_y = 193(4)$ Hz. At the beam splitter duration ($J_y t = 1/8$) bosonic interference leads to a nearly vanishing $\mathcal{P}(1, 1)$, corresponding to an even parity in the output states. This can be interpreted as a measurement of the purity of the initial Fock state, here measured to be 0.90(4). The data shown here are averaged over two independent double wells. The blue curve is a maximum-likelihood fit to the data, and the error bars reflect 1σ statistical error. **c**, When two copies of a product state, such as the Mott insulator in the atomic limit, are interfered on the beam splitter, the output states contain even particle numbers globally (full system) as well as locally (subsystem), indicating pure states in both. **d**, On the other hand, for two copies of an entangled state, such as a superfluid state, the output states contain even particle numbers globally (pure state) but a mixture of odd and even outcomes locally (mixed state). This directly demonstrates entanglement.

lattice potential along x. Then we freeze the tunnelling along x without destroying the coherence in the many-body state and apply the beam splitter along y. Finally, we rapidly turn on a very deep two-dimensional lattice to suppress all tunnelling and detect the atom number parity (even $= 1$, odd $= -1$) at each site. We construct the parity of a spatial region by multiplying the parities of all the sites within that region. The average parity over repeated realizations measures the quantum purity, both globally and locally, according to equation (3), enabling us to determine the second-order Rényi entropy globally and for all possible subsystems.

In the atomic Mott insulator limit (Fig. 3c), the state is separable. Hence, the interference signal between two copies should show even parity in all subsystems, indicating a pure state with zero entanglement entropy. Towards the superfluid regime (Fig. 3d), the build-up of entanglement between various lattice sites leads to mixed states in subsystems, corresponding to a finite entanglement entropy. Hence, the measurement outcomes do not have a pre-determined parity. Remarkably, the outcomes should still retain even global parity, indicating a pure global state. Higher entropy in the subsystems than the global system cannot be explained classically and demonstrates bipartite entanglement.

Experimentally, we find exactly this behaviour for our two 4-site Bose–Hubbard systems (Fig. 4). We observe the emergence of spatial entanglement as the initial atomic Mott insulator melts into a superfluid. The measured quantum purity of the full system is about 0.6 across the Mott insulator to superfluid crossover, corresponding to a Rényi entropy of $S_2(\mathrm{AB}) \approx 0.5$. The measured purity deep in the superfluid phase is slightly reduced, probably owing to the reduced beam splitter fidelity in the presence of increased single-site occupation number, and any residual heating. The nearly constant global purity indicates a high level of coherence throughout the crossover. For lower interaction strength U/J_x (superfluid regime), we observe that the subsystem Rényi entropy is higher than the full system: $S_2(\mathrm{A}) > S_2(\mathrm{AB})$. This demonstrates the presence of spatial entanglement in the superfluid state. In the Mott insulator regime ($U/J_x \gg 1$), $S_2(\mathrm{A})$ is lower than $S_2(\mathrm{AB})$ and proportional to the subsystem size, consistent with a product state.

In these measurements, we post-select outcomes of the experiment for which the total number of atoms detected in both copies is even. This constitutes about 60% of all the data, and excludes realizations with preparation errors, atom loss during the sequence, or detection errors (Supplementary Information). The measured purity is consistent with an imperfect beam splitter operation alone, suggesting much higher purity for the many-body state. The measured entropy is thus a sum of an extensive classical entropy due to the imperfections of the beam splitter and any entanglement entropy.

Our site-resolved measurement simultaneously provides information about all possible spatial partitionings of the system. Comparing the purity of all subsystems with that of the full system enables us to determine whether a quantum state has genuine spatial multipartite entanglement, in which every site is entangled with each other. Experimentally, we find that this is indeed the case for small U/J_x (Fig. 4b). In the superfluid phase, all possible subsystems have more entropy than the full system, demonstrating full spatial multipartite entanglement between all four sites[27,43]. In the Mott phase ($U/J_x \gg 1$), the measured entropy is dominated by extensive classical entropy, showing a lack of entanglement.

By measuring the second-order Rényi entropy we can calculate other useful quantities, such as the associated mutual information $I_{\mathrm{AB}} = S_2(\mathrm{A}) + S_2(\mathrm{B}) - S_2(\mathrm{AB})$. Mutual information exhibits interesting scaling properties with respect to the subsystem size, which can be key to studying area laws in interacting quantum systems[44]. In some cases, such as in 'data hiding states'[45], mutual information is more informative than the more conventional two-point correlators, which might take arbitrarily small values in presence of strong correlations. Mutual information is also immune to extensive classical entropy, and hence has practical utility in the experimental study of larger systems. In our experiments (Fig. 5a), we find that for the Mott insulator state ($U/J_x \gg 1$), the entropy of the full system is the sum of the entropies for the subsystems. The mutual information is $I_{\mathrm{AB}} \approx 0$ for this state, consistent with a product state in the presence of extensive classical entropy. At $U/J_x \approx 10$, correlations between the subsystems begin to grow as the system adiabatically melts into a superfluid, resulting in non-zero mutual information, $I_{\mathrm{AB}} > 0$.

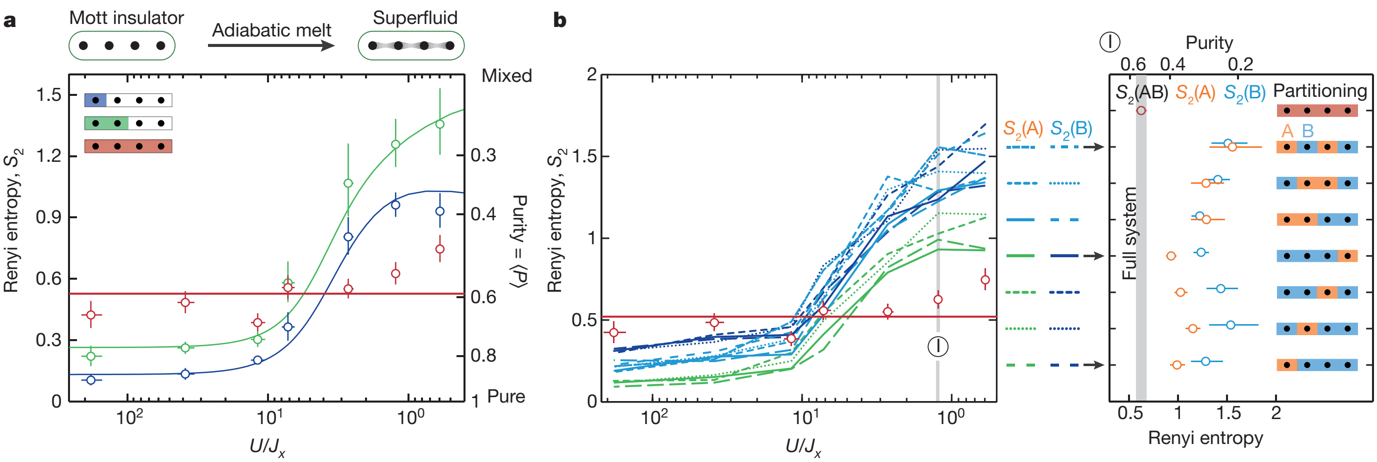


Figure 4 | Entanglement in the ground state of the Bose–Hubbard model. We study the transition from Mott insulator to superfluid with four atoms on four lattice sites in the ground state of the Bose–Hubbard model, equation (4). **a**. As the interaction strength U/J_x is adiabatically reduced, the purity of the subsystem A (green and blue, inset), $\mathrm{Tr}(\rho_A^2)$, becomes less than that of the full system (red). This demonstrates entanglement in the superfluid phase, generated by coherent tunnelling of bosons across lattice sites. In terms of the second-order Rényi entanglement entropy, $S_2(A) = -\log \mathrm{Tr}(\rho_A^2)$, the full system has less entropy than its subsystems in this state. In the Mott insulator phase ($U/J_x \gg 1$) the full system has more Rényi entropy (and less purity) than the subsystems, owing to the lack of sufficient entanglement and a contribution of classical entropy. The circles are data points and the solid lines are theoretical, calculated from exact diagonalization. The only free parameter is an added offset, assumed to be proportional to the system size and consistent with the average measured entropy (about 0.5) in the full system. The vertical error bars in this figure and in Figs 5 and 6 indicate 1σ in combined statistical and systematic errors (Supplementary Information). **b**, Second-order Rényi entropy of all possible bi-partitioning of the system. For small U/J_x, all subsystems (data points connected by green and blue lines) have more entropy than the full system (red circles), indicating full multipartite entanglement[43] between the four lattice sites. The residual entropy in the Mott insulating regime is from classical entropy in the experiment, and extensive in the subsystem size. The right-hand panel in **b** shows the values of all Renyi entropies of the particular case of $U/J_x \approx 1$, to demonstrate spatial multipartite entanglement in this superfluid.

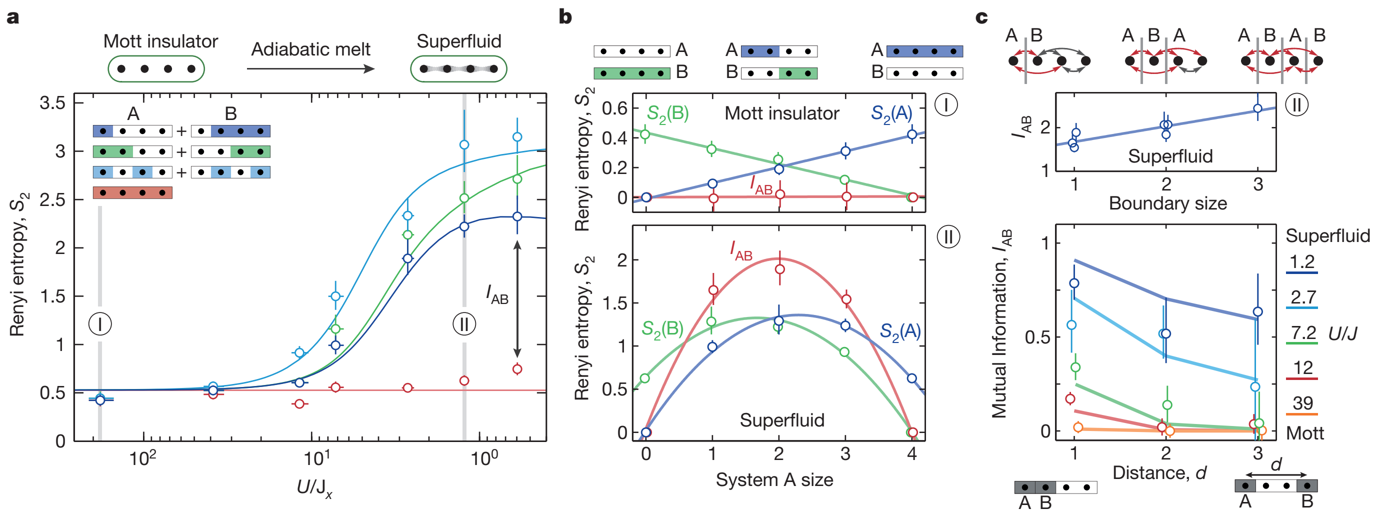


Figure 5 | Rényi mutual information in the ground state. Any contribution from the extensive classical entropy in our measured Rènyi entropy can be factored out by constructing the mutual information $I_{AB} = S_2(A) + S_2(B) - S_2(AB)$. **a**, We plot the summed entropy $S_2(A) + S_2(B)$ (in blue, green and light blue corresponding to the partitions shown) and the entropy of the full system $S_2(AB)$ (in red) separately. Mutual information is the difference between the two, as shown by the arrow for a partitioning scheme. In the Mott insulator phase ($U/J_x \gg 1$) the sites are not correlated, and $I_{AB} \approx 0$. Correlations start to build up for smaller U/J_x, resulting in a non-zero mutual information. The theory curves are from exact diagonalization, with added offsets consistent with the extensive entropy in the Mott insulator phase (about 0.5 for the full system). **b**, Classical and entanglement entropies follow qualitatively different scaling laws in a many-body system. The top panel in **b** shows that in the Mott insulator phase classical entropy dominates and $S_2(A)$ and $S_2(B)$ follow a volume law: entropy increases with the size of the subsystem. The mutual information $I_{AB} \approx 0$. The bottom panel in **b** shows the non-monotonic behaviour of $S_2(A)$ and $S_2(B)$ in the superfluid regime, due to the dominance of entanglement over classical entropy, which makes the curves asymmetric. I_{AB} restores the symmetry by removing the classical uncorrelated noise. The solid lines are linear (top) and quadratic (bottom) fits included as a guide to the eye. The top panel in **c** shows that more correlations are affected (red arrow) with increasing boundary area, leading to a growth of mutual information between subsystems. The data points are for various partitioning schemes shown in Fig. 4b. The bottom panel in **c** plots I_{AB} as a function of the distance d between the subsystems to show the onset and spread of correlations in space, as the Mott insulator adiabatically melts into a superfluid. In these plots some overlapping data points are offset from each other horizontally for clarity.

It is instructive to investigate the scaling of Rényi entropy and mutual information with subsystem size[7,44], since in larger systems they can characterize quantum phases, for example by measuring the central charge of the underlying quantum field theory[5]. Figure 5b shows these quantities versus the subsystem size for various partitioning schemes with a single boundary. For the atomic Mott insulator the Rényi entropy increases linearly with the subsystem size and the mutual information is zero, consistent with both a product state and classical entropy being uncorrelated between various sites. In the superfluid state the measured Rényi entropy curves are asymmetric and first increase with the system size, then fall again as the subsystem size approaches that of the full system. This represents the combination of entanglement entropy and the linear classical entropy. The non-monotonicity is a signature of the entanglement entropy, as the entropy for a pure state must vanish when the subsystem size is zero or the full system. The asymmetry due to classical entropy is absent in the mutual information.

The mutual information between two subsystems comes from the correlations across their separating boundary. For a 4-site system, the boundary size ranges from one to three for various partitioning schemes. Among those schemes with a single boundary, maximum mutual information in the superfluid is obtained when the boundary divides the system symmetrically (Fig. 5a). Increasing the boundary size increases the mutual information, as more correlations are interrupted by the partitioning (Fig. 5c).

Mutual information also elucidates the onset of correlations between various sites as the few-body system crosses over from a Mott insulator to a superfluid phase. In the Mott insulator phase ($U/J_x \gg 1$) the mutual information between all sites vanish (Fig. 5c, bottom). As the particles start to tunnel, only the nearest-neighbour correlations start to build up ($U/J_x \approx 12$) and the long-range correlations remain negligible. Further into the superfluid phase, the correlations extend beyond the nearest neighbour and become long range for smaller U/J_x. These results suggest disparate spatial behaviour of the mutual information in the ground state of an uncorrelated (Mott insulator) and a strongly correlated phase (superfluid). For larger systems this can be exploited to identify quantum phases and the onset of quantum phase transitions.

Non-equilibrium entanglement dynamics

Away from the ground state, the non-equilibrium dynamics of a quantum many-body system is often theoretically intractable. This is due to the growth of entanglement beyond the access of numerical techniques, such as the time-dependent density matrix renormalization group theory[46,47]. Experimental investigation of entanglement may shed valuable light onto non-equilibrium quantum dynamics. Towards this goal, we study a simple system: two particles oscillating in a double well[37,48]. The non-equilibrium dynamics are described by the Bose–Hubbard model. The quantum state of the system oscillates between unentangled (particles localized in separate wells) states and entangled states in the Hilbert space spanned by $|1, 1\rangle$, $|2, 0\rangle$ and $|0, 2\rangle$. Here, $|m, n\rangle$ denotes a state with m and n atoms in the two subsystems (wells), respectively. Starting from the product state $|1, 1\rangle$ the system evolves through the maximally entangled states $|2, 0\rangle + |0, 2\rangle \pm |1, 1\rangle$ and the symmetric, HOM-like state $|2, 0\rangle + |0, 2\rangle$. In the maximally entangled states the subsystems are completely mixed, with a probability of 1/3 of having zero, one or two particles. The system then returns to the initial product state $|1, 1\rangle$ before re-entangling. In our experiment, we start with a Mott insulating state ($U/J_x \gg 1$), and suddenly quench the interaction parameter to a low value, $U/J_x \approx 0.3$. The non-equilibrium dynamics is demonstrated (Fig. 6) by the oscillation in the second-order Rényi entropy of the subsystem, while the full system assumes a constant value originating from classical entropy. This experiment also demonstrates entanglement in HOM-like interference of two massive particles.

Summary and outlook

In this work, we perform a direct measurement of quantum purity, the second-order Rényi entanglement entropy, and mutual information

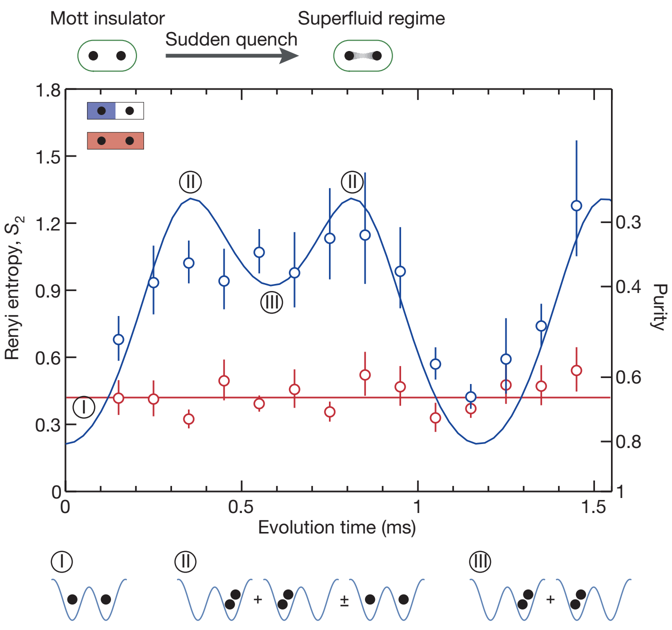


Figure 6 | Entanglement dynamics in a quench. Entanglement dynamics of two atoms in two sites after a sudden quench of the Hamiltonian from a large value of U/J_x to $U/J_x \approx 0.3$, with $J_x \approx 210$ Hz. Here, 'evolution time' refers to the duration that the atoms spend in the shallow double well, after the initial sudden quench. The system oscillates between Mott insulator (I) and quenched superfluid regimes (II, III). The growth of bipartite entanglement in the superfluid regime is seen by comparing the measured Rényi entropy of the single site subsystem (blue data points) to that of the two site full system (red data points). The solid lines are the theoretical curves, with vertical offsets to include the classical entropy introduced by experimental imperfections.

in a Bose–Hubbard system. Our measurement scheme does not rely on full density matrix reconstruction or the use of specialized witness operators to detect entanglement. Instead, by preparing and interfering two identical copies of a many-body quantum state, we probe entanglement with the measurement of only a single operator. Our experiments represent an important demonstration of the usefulness of the many-body interference for the measurement of entanglement. It is straightforward to extend the scheme to fermionic systems[49] and systems with internal degrees of freedom[27], and to two dimensions. By generalizing the interference to n copies of the quantum state[29], arbitrary observables written as an nth-order polynomial function of the density matrix—for example, Rényi entropies of order $n > 2$—can be measured.

With modest technical upgrades to suppress classical fluctuations and residual interactions, it should be possible to further improve the beam splitter fidelity, enabling us to work with much larger systems. Mutual information may be ideal for exploring larger systems as it is insensitive to any residual extensive classical entropy. For high entropy of a subsystem, corresponding to low state purity, the number of measurements required to reach a desired precision is high. However, in contrast to tomographic methods, this scheme would not require additional operations for larger systems. Moreover, the single-site resolution of the microscope allows us to simultaneously obtain information about all possible subsystems, to probe multipartite entanglement.

For non-equilibrium systems, entanglement entropy can grow in time (indefinitely in infinite systems). This leads to interesting many-body physics, such as thermalization in closed quantum systems[50]. The long duration of growth of entanglement entropy is considered to be a key signature of many-body localized states[14] arising in the presence of disorder. The ability to measure the quantum purity for these systems would allow experimental distinction of quantum fluctuations and classical statistical fluctuations.

More generally, by starting with two different quantum states in the two copies this scheme can be applied to measure the quantum state overlap between them. This would provide valuable information about the underlying quantum state. For example, the many-body ground state is very sensitive to perturbations near a quantum critical point. Hence, the overlap between two ground states with slightly different parameters (such as U/J in the Bose–Hubbard Hamiltonian) could be used as a sensitive probe of quantum criticality[51]. Similarly the overlap of two copies undergoing non-equilibrium evolution under different perturbations can be used to probe temporal correlation functions in non-equilibrium quantum dynamics.

Received 30 July; accepted 16 September 2015.

1. Horodecki, R., Horodecki, P., Horodecki, M. & Horodecki, K. Quantum entanglement. *Rev. Mod. Phys.* **81,** 865 (2009).
2. Aspect, A. Bell's inequality test: more ideal than ever. *Nature* **398,** 189–190 (1999).
3. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2010).
4. Amico, L., Fazio, R., Osterloh, A. & Vedral, V. Entanglement in many-body systems. *Rev. Mod. Phys.* **80,** 517 (2008).
5. Calabrese, P. & Cardy, J. Entanglement entropy and conformal field theory. *J. Phys. A* **42,** 504005 (2009).
6. Nishioka, T., Ryu, S. & Takayanagi, T. Holographic entanglement entropy: an overview. *J. Phys. A* **42,** 504008 (2009).
7. Eisert, J., Cramer, M. & Plenio, M. B. Colloquium: area laws for the entanglement entropy. *Rev. Mod. Phys.* **82,** 277 (2010).
8. Kitaev, A. & Preskill, J. Topological entanglement entropy. *Phys. Rev. Lett.* **96,** 110404 (2006).
9. Levin, M. & Wen, X.-G. Detecting topological order in a ground state wave function. *Phys. Rev. Lett.* **96,** 110405 (2006).
10. Jiang, H.-C., Wang, Z. & Balents, L. Identifying topological order by entanglement entropy. *Nature Phys.* **8,** 902–905 (2012).
11. Zhang, Y., Grover, T. & Vishwanath, A. Entanglement entropy of critical spin liquids. *Phys. Rev. Lett.* **107,** 067202 (2011).
12. Isakov, S. V., Hastings, M. B. & Melko, R. G. Topological entanglement entropy of a Bose–Hubbard spin liquid. *Nature Phys.* **7,** 772–775 (2011).
13. Vidal, G., Latorre, J. I., Rico, E. & Kitaev, A. Entanglement in quantum critical phenomena. *Phys. Rev. Lett.* **90,** 227902 (2003).
14. Bardarson, J. H., Pollmann, F. & Moore, J. E. Unbounded growth of entanglement in models of many-body localization. *Phys. Rev. Lett.* **109,** 017202 (2012).
15. Daley, A. J., Pichler, H., Schachenmayer, J. & Zoller, P. Measuring entanglement growth in quench dynamics of bosons in an optical lattice. *Phys. Rev. Lett.* **109,** 020505 (2012).
16. Schuch, N., Wolf, M. M., Verstraete, F. & Cirac, J. I. Entropy scaling and simulability by matrix product states. *Phys. Rev. Lett.* **100,** 030504 (2008).
17. Bloch, I., Dalibard, J. & Nascimbène, S. Quantum simulations with ultracold quantum gases. *Nature Phys.* **8,** 267–276 (2012).
18. Blatt, R. & Roos, C. F. Quantum simulations with trapped ions. *Nature Phys.* **8,** 277–284 (2012).
19. Aspuru-Guzik, A. & Walther, P. Photonic quantum simulators. *Nature Phys.* **8,** 285–291 (2012).
20. Houck, A. A., Türeci, H. E. & Koch, J. On-chip quantum simulation with superconducting circuits. *Nature Phys.* **8,** 292–299 (2012).
21. Bouwmeester, D., Pan, J.-W., Daniell, M., Weinfurter, H. & Zeilinger, A. Observation of three-photon Greenberger-Horne-Zeilinger entanglement. *Phys. Rev. Lett.* **82,** 1345–1349 (1999).
22. Gühne, O. & Tóth, G. Entanglement detection. *Phys. Rep.* **474,** 1–75 (2009).
23. James, D. F. V., Kwiat, P. G., Munro, W. J. & White, A. G. Measurement of qubits. *Phys. Rev. A* **64,** 052312 (2001).
24. Pan, J.-W. *et al.* Multiphoton entanglement and interferometry. *Rev. Mod. Phys.* **84,** 777 (2012).
25. Häffner, H. *et al.* Scalable multiparticle entanglement of trapped ions. *Nature* **438,** 643–646 (2005).
26. Ekert, A. K. *et al.* Direct estimations of linear and nonlinear functionals of a quantum state. *Phys. Rev. Lett.* **88,** 217901 (2002).
27. Moura Alves, C. & Jaksch, D. Multipartite entanglement detection in bosons. *Phys. Rev. Lett.* **93,** 110501 (2004).
28. Bakr, W. S. *et al.* Probing the superfluid–to–Mott insulator transition at the single-atom level. *Science* **329,** 547–550 (2010).
29. Brun, T. A. Measuring polynomial functions of states. *Quantum Inform. Comput.* **4,** 401–408 (2004).
30. Bovino, F. A. *et al.* Direct measurement of nonlinear properties of bipartite quantum states. *Phys. Rev. Lett.* **95,** 240407 (2005).
31. Walborn, S. P., Ribeiro, P. S., Davidovich, L., Mintert, F. & Buchleitner, A. Experimental determination of entanglement with a single measurement. *Nature* **440,** 1022–1024 (2006).
32. Schmid, C. *et al.* Experimental direct observation of mixed state entanglement. *Phys. Rev. Lett.* **101,** 260505 (2008).

33. Horodecki, R. & Horodecki, M. Information-theoretic aspects of inseparability of mixed states. *Phys. Rev. A* **54,** 1838 (1996).
34. Mintert, F. & Buchleitner, A. Observable entanglement measure for mixed quantum states. *Phys. Rev. Lett.* **98,** 140505 (2007).
35. Li, H. & Haldane, F. D. M. Entanglement spectrum as a generalization of entanglement entropy: identification of topological order in non-abelian fractional quantum Hall effect states. *Phys. Rev. Lett.* **101,** 010504 (2008).
36. Hong, C. K., Ou, Z. Y. & Mandel, L. Measurement of subpicosecond time intervals between two photons by interference. *Phys. Rev. Lett.* **59,** 2044 (1987).
37. Kaufman, A. M. *et al.* Two-particle quantum interference in tunnel-coupled optical tweezers. *Science* **345,** 306–309 (2014).
38. Lopes, R. *et al.* Atomic Hong–Ou–Mandel experiment. *Nature* **520,** 66–68 (2015).
39. Cramer, M. *et al.* Spatial entanglement of bosons in optical lattices. *Nat. Commun.* **4,** 2161 (2013).
40. Fukuhara, T. *et al.* Spatially resolved detection of a spin-entanglement wave in a Bose–Hubbard chain. *Phys. Rev. Lett.* **115,** 035302 (2015).
41. Bartlett, S. D. & Wiseman, H. M. Entanglement constrained by superselection rules. *Phys. Rev. Lett.* **91,** 097903 (2003).
42. Schuch, N., Verstraete, F. & Cirac, J. I. Nonlocal resources in the presence of superselection rules. *Phys. Rev. Lett.* **92,** 087904 (2004).
43. Palmer, R. N., Moura Alves, C. & Jaksch, D. Detection and characterization of multipartite entanglement in optical lattices. *Phys. Rev. A* **72,** 042335 (2005).
44. Wolf, M. M., Verstraete, F., Hastings, M. B. & Cirac, J. I. Area laws in quantum systems: mutual information and correlations. *Phys. Rev. Lett.* **100,** 070502 (2008).
45. Terhal, B. M., DiVincenzo, D. P. & Leung, D. W. Hiding bits in Bell states. *Phys. Rev. Lett.* **86,** 5807–5810 (2001).
46. Vidal, G. Efficient simulation of one-dimensional quantum many-body systems. *Phys. Rev. Lett.* **93,** 040502 (2004).
47. Trotzky, S. *et al.* Probing the relaxation towards equilibrium in an isolated strongly correlated one-dimensional Bose gas. *Nature Phys*. **8,** 325–330 (2012).
48. Trotzky, S., Chen, Y.-A., Schnorrberger, U., Cheinet, P. & Bloch, I. Controlling and detecting spin correlations of ultracold atoms in optical lattices. *Phys. Rev. Lett.* **105,** 265303 (2010).
49. Pichler, H., Bonnes, L., Daley, A. J., Läuchli, A. M. & Zoller, P. Thermal versus entanglement entropy: a measurement protocol for fermionic atoms with a quantum gas microscope. *New J. Phys.* **15,** 063003 (2013).
50. Rigol, M., Dunjko, V. & Olshanii, M. Thermalization and its mechanism | for generic isolated quantum systems. *Nature* **452,** 854–858 (2008).
51. Zanardi, P. & Paunković, N. Ground state overlap and quantum phase transitions. *Phys. Rev. E* **74,** 031123 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Abanin, J. I. Cirac, M. Cramer, A. Daley, A. DelMaestro, E. Demler, M. Endres, S. Gopalakrishnan, M. Headrick, A. Kaufman, M. Knap, T. Monz, A. Pal, H. Pichler, S. Sachdev, B. Swingle, P. Zoller, and M. Zwierlein for useful discussions. This work was supported by grants from the Gordon and Betty Moore Foundations EPiQS Initiative (grant GBMF3795), the NSF through the Center for Ultracold Atoms, the Army Research Office with funding from the DARPA OLE programme and a MURI programme, an Air Force Office of Scientific Research MURI programme, and an NSF Graduate Research Fellowship (to M.R.).

Author Contributions All authors contributed to the construction and execution of the experiments, data analysis and the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.G. (greiner@physics.harvard.edu).

doi:10.1038/nature15736

Pharmacogenomic agreement between two cancer cell line data sets

The Cancer Cell Line Encyclopedia and Genomics of Drug Sensitivity in Cancer Investigators*

Large cancer cell line collections broadly capture the genomic diversity of human cancers and provide valuable insight into anti-cancer drug response. Here we show substantial agreement and biological consilience between drug sensitivity measurements and their associated genomic predictors from two publicly available large-scale pharmacogenomics resources: The Cancer Cell Line Encyclopedia and the Genomics of Drug Sensitivity in Cancer databases.

Performing *in vitro* pharmacological sensitivity studies across panels of molecularly characterized cancer cell lines has proved useful in assessing the cellular activity of many compounds, assigning mechanisms of drug action, and determining genetic contexts for distinct cancer vulnerabilities[1–6]. A recent comparison study[7] of the Cancer Cell Line Encyclopedia (CCLE)[8] and the Genomics of Drug Sensitivity in Cancer (GDSC)[9] reported poor correlations between their pharmacological data, thus questioning the validity of some conclusions. These observations raised important questions for the field about how best to perform comparisons of large-scale data sets, evaluate the robustness of such studies, and interpret their analytical outputs.

To address these questions, we performed a systematic comparison of the CCLE and GDSC pharmacological data and drug sensitivity predictors. Our results show that when biologically-grounded analytical considerations are incorporated, pharmacological data from the CCLE and GDSC studies exhibit reasonable consistency. Most importantly, these analyses demonstrate that data from either study yields similar predictors of drug response.

Comparison of cell line pharmacological data sets

To evaluate the consistency of the pharmacological data from the two studies, we first performed a comparative analysis of CCLE and GDSC drug screening metrics. For this analysis, we used both the 50% inhibitory concentration (IC_{50}) and the area under the curve (AUC; also referred to as activity area in CCLE when considering $1-$AUC). Importantly, IC_{50} values were capped at the maximum tested drug concentrations to ensure that they could be properly compared between both data sets (Supplementary Data 1). Also, a fixed axis scale was applied across all compounds to facilitate visualization (Extended Data Fig. 1). Of note, while 471 cell lines are present in both CCLE and GDSC collections and have associated genomic data, only a subset of those have overlapping drug screening data: a range of 82–256 cell lines per compound (median = 94 cell lines; mean = 157; Fig. 1a and Supplementary Data 1).

Our analytical approach was designed to account for the fact that many pharmacological profiles exhibit highly discontinuous distributions across cancer cell line collections. Whereas a subset of individual lines may show marked pharmacological sensitivity, the remaining lines—often the vast majority of cell lines in the collection—may be relatively insensitive to a given drug. Such 'outlier' distributions are expected, as they are typically observed for drugs that target specific oncogenic dependencies. Given the relative paucity of sensitive outliers, appropriate pharmacological assessments require multiple drug-sensitive cell lines for each compound and the ability to discern this relevant signal against a background dominated by the insensitive majority. Additionally, small data sets containing exclusively insensitive lines are not expected to display significant correlations given the inherent noise in their drug response data.

In cases where direct GDSC–CCLE comparisons were possible, nearly all compounds (13/15) exhibited AUC and IC_{50} distributions dominated by drug-insensitive lines, with a much smaller number of drug-sensitive outliers. The complete distributions of all CCLE and GDSC AUC values are illustrated for each compound by "'violin plots", while overlapping lines are displayed as a scatter plot (representative examples are shown in Fig. 1, and all plots in Extended Data Fig. 1); results for IC_{50} values are similar (Extended Data Fig. 1). Ten compounds (saracatinib (also known as AZD0530), erlotinib, lapatinib, nilotinib, crizotinib, nutlin-3, PD0332991, PHA665752, PLX4720 and sorafenib) exhibited AUC values skewed heavily towards the drug-insensitive end of the spectrum. Notably, several targeted anticancer drugs had very few (if any) drug-sensitive lines in the overlapping set (for example, 2 for crizotinib, 3 for nilotinib, 2 for TAE684, and zero for erlotinib or sorafenib; Fig. 1b, c and Extended Data Fig. 1). This relative paucity of drug-sensitive cell lines in the overlapping set constrained the level of correlation achievable.

Nevertheless, a correlation analysis that accounted for the imbalance between the number of sensitive and insensitive cell lines, and for differences in the original analytical methodologies, yielded good consistency in most cases (see Extended Data Fig. 2, comparing Spearman's and Pearson's correlations properties in this context, and Supplementary Discussion). When using the Pearson correlation coefficient instead of Spearman's, as well as consistently capped drug sensitivity metrics, correlation values were clearly improved for most drugs compared to the earlier comparison study[7] (Fig. 1d, e, Methods and Supplementary Discussion). We noted that some correlation values remained poor, either owing to differences in cell line biology, in actual pharmacological measurements (for example, nutlin-3, paclitaxel and PHA665752), or because sensitive lines were only present in one of the cell line collections (for example, erlotinib and sorafenib), preventing any meaningful comparison (Fig. 1c).

To complement this correlation analysis, we used a waterfall plot-based assessment (Extended Data Fig. 3 shows a schematic of the workflow and further details are provided in the Supplementary Discussion). This analysis confirmed that on average, 94% of cell lines for the 13 relevant compounds (CCLE mean = 94%, range = 77–100%; GDSC mean = 96%, range = 86–100%; Supplementary Data 2) clustered within a drug-insensitive range (for example, IC_{50} values of $>1\,\mu$M for most compounds). These waterfall analyses also showed a high consistency of

*Lists of participants and their affiliations appear at the end of the paper.

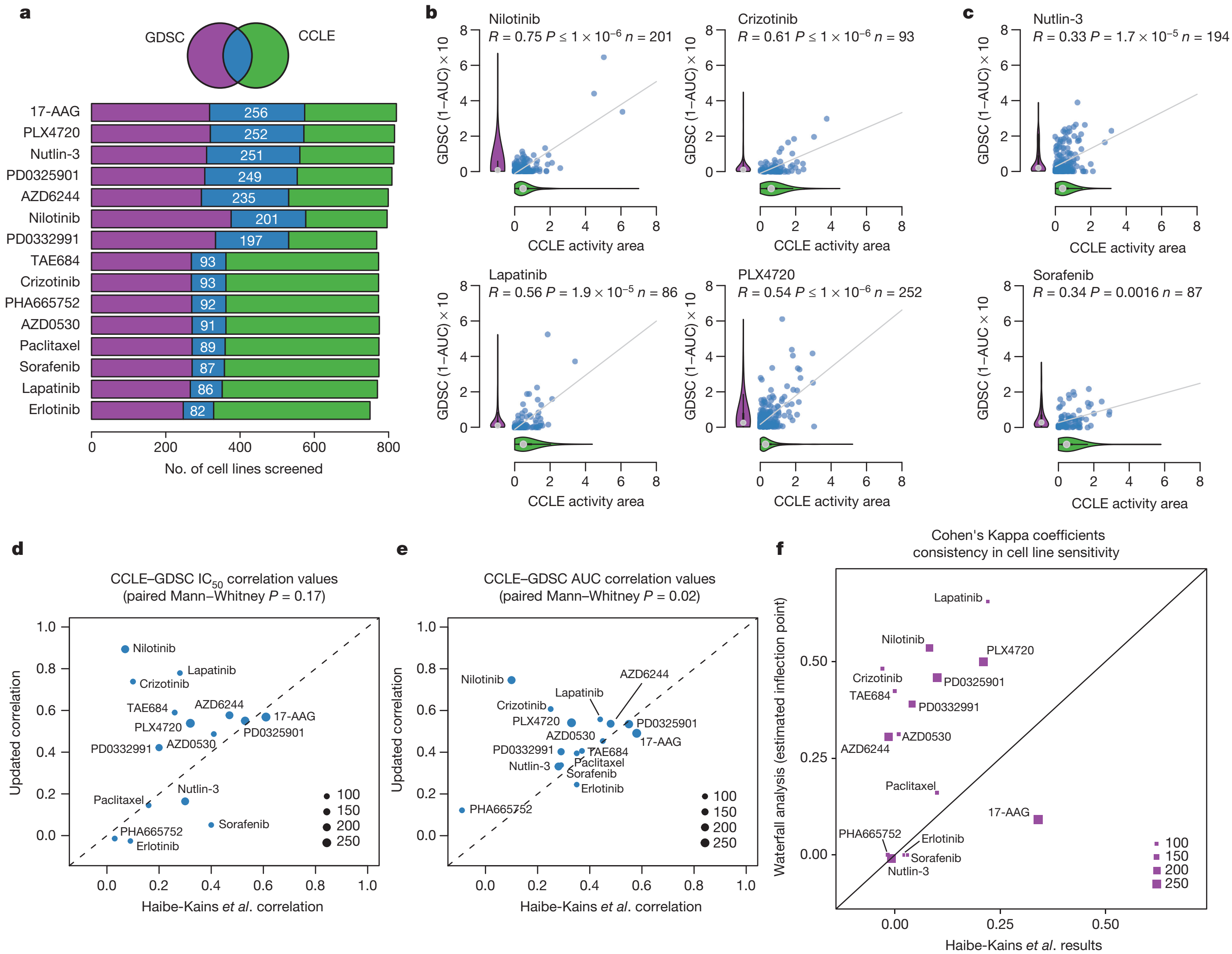


Figure 1 | Comparison of pharmacological data from the CCLE and GDSC studies. **a**, Overlap of data sets. **b**, **c**, Comparison of drug sensitivity (AUC) measured in *n* overlapping cell lines between the studies for drugs with good (**b**) or poor (**c**) correlation. *R*, Pearson correlation coefficient; *P*, *P* value. Violin plots, distribution of sensitivity values for all lines in each study. Grey dot, median; black line, interquartile range; shape, kernel density of the distribution. **d**, **e**, Correlation coefficients between GDSC and CCLE data sets. *x* axis, Spearman, Haibe-Kains *et al.*[7]; *y* axis, Pearson, present analysis. Dot sizes are proportional to the number of overlapping cell lines. Dots above the dashed $y = x$ line denote an improved correlation compared to Haibe-Kains *et al.*[7]. **f**, Comparisons of Cohen's Kappa coefficient testing studies' agreement in Haibe-Kains *et al.*[7] (*x* axis) and the present study (*y* axis) for sensitivity/resistance calling using a waterfall plot analysis.

cell line categorization as "sensitive" or "resistant" between CCLE and GDSC data (Fig. 1f, Extended Data Fig. 3). This consistency was evident even when using a simple drug sensitivity cut-off (1 μM) across all the drugs tested (Extended Data Fig. 3). Thus, both categorization approaches showed higher consistency than reported in the earlier study[7] (see Supplementary Discussion). These results indicated that the CCLE and GDSC cell line pharmacological screening data are best suited for modelling studies that distinguish rare, drug-sensitive lines from "all others" (for example, from drug-insensitive lines that are not expected to contribute meaningful molecular or genetic information).

Comparison of drug sensitivity predictors

Next, we considered the extent to which the CCLE and GDSC cell line collections illuminated common genetic or molecular underpinnings of anticancer drug efficacy. Such insights provide one of the most relevant measures for concordance and utility of pharmacological screening data, given that these efforts are designed to identify such predictors of drug response. First, we determined whether molecular correlates of drug response were aligned between the two data sets. Here we performed an analysis of variance (ANOVA) using the overlapping lines across the CCLE and GDSC. We considered two models where the predicted variables were IC_{50} values or activity area (that is, 1 − AUC) scores, respectively. In both models we considered the tissue of origin as a covariate and the mutational status of 71 oncogenes as independent variables.

ANOVA identified known genetic biomarkers of sensitivity or resistance as top molecular correlates in at least one data set for 13/15 compounds, and in both data sets for 8/15 compounds (Fig. 2a, Extended Data Fig. 4, Supplementary Data 3). Genetic correlates in both data sets included *NRAS* mutation and sensitivity to MEK inhibitor PD0325901, *BRAF* mutations and sensitivity to BRAF inhibitor PLX4720, the *BCR-ABL1* fusion gene and sensitivity to multiple ABL1 inhibitors (nilotinib, AZD0530) and sensitivity of *ERBB2*-amplified cells to ERBB2 inhibitor lapatinib (identified when using IC_{50} values; Extended Data Fig. 4). Additionally, drug resistance associations such as *TP53* mutations and resistance to nutlin-3 were recovered consistently using activity area scores. When ANOVA was fitted to activity area, 14 drugs for the GDSC and 15 for the CCLE also showed lineage-specific response associations that were consistent across data sets (post-hoc Welch *t*-test; Extended Data Fig. 5 and Supplementary Data 4 and 7).

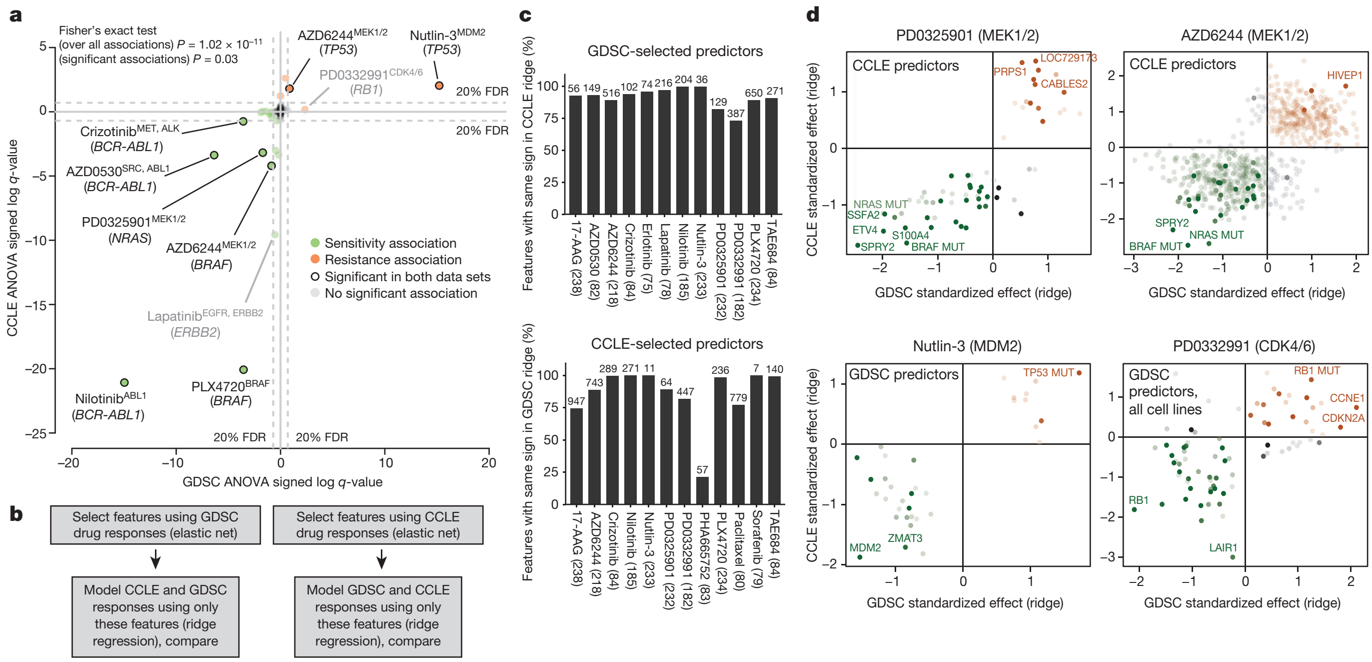


Figure 2 | Consistency of drug sensitivity prediction markers between the CCLE and GDSC data sets. a, ANOVA on overlapping data set (1 − AUC). Coordinates, 'signed log *q*-values'. Negative sign, gene associated with increased sensitivity; positive, increased resistance. Distance from 0, *q*-value. Fisher's exact test of consistency of marker behaviour on all or only significant associations. Markers in grey are not significant; markers highlighted are significant in both the studies. Corresponding drug name, target(s) and cancer gene are reported for a subset of therapeutically relevant interactions. FDR, false discovery rate. **b**–**d**, Elastic net and ridge regression analysis. **b**, Analytical strategy. **c**, Proportion of genomic features with consistent effect on drug response in both studies (total number of features tested displayed above the bar and number cell lines indicated in parentheses). **d**, Ridge regression using predictors selected by elastic net. Contrast, frequency of selection in 100 independent elastic net runs. Green and red, association with sensitivity or resistance, respectively.

In a more comprehensive assessment of the consistency of genomic predictors, we applied a multivariate analysis across 21,013 genomic features encompassing expression, copy number changes and mutations[8,9]. Elastic net regression was performed using either the full data set available for each study or only the overlapping data sets. This analysis yielded robust response predictors, and the overlap of predictors was highly significant (χ^2 $P < 10^{-8}$; Extended Data Fig. 6, Supplementary Data 5). Here again, known genomic predictors of drug response emerged as top molecular correlates in at least one data set for 13/15 compounds; 10/15 compounds showed such correlates in both data sets (Supplementary Data 5), as reported previously by CCLE and GDSC using their individual data sets[8,9]. For some drugs, extending elastic net regression analyses of IC_{50} values beyond just the overlapping cell lines identified additional genetic predictors of clinical activity. *MDM2* expression and *TP53* mutations in the case of nutlin-3 sensitivity provide one example. Moreover, among 4,957 drug–gene associations found using elastic net modelling on each data set, we only observed one divergent result (0.02%) between the two studies.

To further explore how the two data sets might be leveraged to identify genomic predictors of drug sensitivity, we performed a two-step analysis where predictors were identified using one data set and their effects were analysed in the other data set. Here, we used elastic net regression to identify the genomic features and ridge regression to compare their effect across the data sets (Fig. 2b and Supplementary Discussion). Additionally, we performed this discovery step either on the overlapping cell lines or on all lines available in the respective studies.

We again observed a high consistency of predictive genomic features identified across the CCLE and GDSC studies, even for drugs where few overlapping cell lines were available. Indeed, > 80% of these features identified with concordant directionality in both studies (Fig. 2c, d, Extended Data Figs 7–9 and Supplementary Data 6, features with same sign). In some instances, no predictors could be identified by the initial elastic net regression. This was often attributable at least in part to small numbers of drug-sensitive cell lines, as noted above (Extended Data Fig. 10). On the other hand, some drugs that exhibited low correlations based on the AUC or IC_{50} analyses nonetheless enabled identification of consistent predictors (for example, nutlin-3; Fig. 2d).

Together, these results indicate that the CCLE and GDSC pharmacological data sets exhibit reasonable predictive power both separately and when taken as a whole. Many of the resulting drug response predictions are well validated by prior knowledge and clinical evidence. In this regard, not only do the two sets of drug screening data exhibit broad convergence—they also provide examples of consilience: a phenomenon in which independent lines of experimental evidence, each with their own inherent limitations, arrive at fundamental scientific agreement.

Discussion

In summary, when analytical and biological considerations are incorporated that reflect the nature of oncogenic dependency, pharmacological data from the CCLE and GDSC studies exhibit reasonable consistency. Based on positive Pearson correlations ($R > 0.5$), we observed agreement across the CCLE and GDSC data sets for the majority (67%) of evaluable compounds (two drugs with clear positive regression slopes showed R values just under 0.5 for the IC_{50} values; Extended Data Fig. 1). We acknowledge that the consistency is not perfect: numerous biological and methodological components (for example, numbers of cell lines seeded per well, drug concentration range examined, number of cell doublings achieved, cell viability assays, analytical tools to calculate sensitivity values, and so on) undoubtedly reduced the statistical correlation of the overlapping pharmacological data. Further standardization of such methodologies will certainly improve correlation metrics, and we welcome efforts in this direction. Nonetheless, both the CCLE and GDSC groups used standard methods for testing drug responses in cell lines, and this analysis confirmed that the consistency of their results seems reasonable in light of the aforementioned methodological differences.

The identification of molecular predictors of drug response remains a major challenge for cancer precision medicine. Accordingly, large-scale

screening of clinically relevant compounds across molecularly annotated cancer cell line collections is likely to remain a crucial preclinical source for hypothesis generation. The CCLE[8] and GDSC[9] data sets, the two biggest public collections of genomic and pharmacological cell line data, have produced largely concordant results thus far, although rigorous comparisons should continue to be performed as these data sets evolve. Although neither data set is perfect on its own, they have both shown clear utility for predictive modelling studies and, in several cases, convergence onto known biological principles. Principled analytical frameworks (together with improved standardization) may conceivably illuminate additional areas of consilience through comparative studies of other functional screens (for example, RNA interference, CRISPR genome editing, phospho-proteomics, etc.) in the future. In all such instances, knowledge of the underlying biology should guide the implementation of those analytical and statistical methods best suited for comparative studies and, more generally, the extraction of meaning from large-scale screening data in cancer and other disease models.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 March 2014; accepted 15 September 2015.
Published online 16 November 2015.

1. Sharma, S. V., Haber, D. A. & Settleman, J. Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. *Nature Rev. Cancer* **10,** 241–253 (2010).
2. Neve, R. M. *et al.* A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10,** 515–527 (2006).
3. Caponigro, G. & Sellers, W. R. Advances in the preclinical testing of cancer therapeutic hypotheses. *Nature Rev. Drug Discov.* **10,** 179–187 (2011).
4. Garraway, L. A. *et al.* Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436,** 117–122 (2005).
5. Solit, D. B. *et al.* BRAF mutation predicts sensitivity to MEK inhibition. *Nature* **439,** 358–362 (2006).
6. Sos, M. L. *et al.* Predicting drug susceptibility of non-small cell lung cancers based on genetic lesions. *J. Clin. Invest.* **119,** 1727–1740 (2009).
7. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504,** 389–393 (2013).
8. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–607 (2012).
9. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483,** 570–575 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. Golub, E. Lander, S. Schreiber, P. Clemons and J. Engelman for helpful discussions. This work was supported by research grants from the Novartis Institutes for BioMedical Research (CCLE; L.A.G., M.G., and G.V.K.) and by grants from the Wellcome Trust (086357 and 102696; D.A.H., M.R.S., U.M., M.J.G., A.A., C.H.B.) and the National Institutes of Health (1U54HG006097-01, A.A. and C.H.B.). L.A.G. was supported in part by grants from Novartis and the Dr. Miriam and Sheldon Adelson Medical Research Foundation. G.V.K. was supported in part by the Slim Foundation. F.I. was supported in part by the EMBL-EBI and Wellcome Trust Sanger Institute Post-Doctoral (ESPOD) programme, and U.M. was funded by a Cancer Research UK Clinician Scientist Fellowship (A16629).

Author Contributions N.S., L.A.G., A.A., D.A.H., C.H.B., S.R., J.L., J.B., G.C., R.S., W.R.S., F.S., M.P.M., F.I., M.M., J.S.-R., M.R.S., U.M. and M.J.G. conceived the studies; N.S., M.G., G.V.K., A.A., I.P.-M., J.L., M.L., D.S., A.K., K.V., E.J.E., M.P.M., F.I. and M.M. performed analyses; N.S., M.G., A.A., M.L., M.R., F.I. and M.M. wrote/tested the R code, and N.S., M.G., A.A., L.A.G., C.H.B., J.L., M.P.M. and J.S.-R. wrote the paper. The Cancer Cell Line Encyclopedia investigators are N.S., M.G., G.V.K., L.A.G., J.L., M.L., D.S., A.K., K.V., E.J.E., M.R., J.B., G.C., R.S., W.R.S., F.S. and M.M.; the Genomics of Drug Sensitivity in Cancer investigators are A.A., I.P.-M., D.A.H., S.R., C.H.B., F.I., M.R.S., U.M. and M.J.G.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.A.G. (Levi_Garraway@dfci.harvard.edu) or C.H.B. (CBENES@mgh.harvard.edu).

The Cancer Cell Line Encyclopedia Consortium

Broad Institute Nicolas Stransky[1*†], Mahmoud Ghandi[1*], Gregory V. Kryukov[1], Levi A. Garraway[1,2]; **Novartis Institutes for Biomedical Research** Joseph Lehár[4*†], Manway Liu[4†], Dmitriy Sonkin[4†], Audrey Kauffmann[4], Kavitha Venkatesan[4], Elena J. Edelman[4], Markus Riester[4], Jordi Barretina[4], Giordano Caponigro[4], Robert Schlegel[4], William R. Sellers[4], Frank Stegmeier[4] & Michael Morrissey[4]

The Genomics of Drug Sensitivity in Cancer Consortium

Massachusetts General Hospital Arnaud Amzallag[3*], Iulian Pruteanu-Malinici[3], Daniel A. Haber[3], Sridhar Ramaswamy[3], Cyril H. Benes[3]; **European Molecular Biology Laboratory, European Bioinformatics Institute, and Wellcome Trust Sanger Institute** Michael P. Menden[5*], Francesco Iorio[5,6*], Michael R. Stratton[6], Ultan McDermott[6], Mathew J. Garnett[6] & Julio Saez-Rodriguez[5†]

[1]The Broad Institute of MIT and Harvard, 415 Main Street, Cambridge, Massachusetts 02142, USA. [2]Department of Medical Oncology, Dana-Farber Cancer Institute, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Massachusetts General Hospital Cancer Center, 149 13th Street, Charlestown, Massachusetts 02129, USA. [4]Novartis Institutes for BioMedical Research, 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. [5]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton CB10 1SD, UK. [6]Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. †Present addresses: Blueprint Medicines, 38 Sidney Street, Cambridge, Massachusetts 02139, USA (N.S.); Google Life Sciences, 1600 Amphitheatre Pkwy, Mountain View, California 94043, USA (J.L. and M.L.); National Cancer Institute, 9609 Medical Center Drive, Rockville, Maryland 20850, USA (D.S.); Joint Research Centre for Computational Biomedicine, RWTH Aachen University, Faculty of Medicine, D-52074 Aachen, Germany (J.S.-R.).
*These authors contributed equally to this work.

METHODS

Power analysis. To estimate and compare the statistical power of Spearman and Pearson correlation tests, we ran the following simulation using synthetically generated drug data: starting with 1,000 cell lines with α percent of them drug-sensitive ($\alpha = 2\%$, 5%, 10%, and 50%), we randomly selected a subset of N samples (N between 3 and 500) and calculated the Spearman and Pearson correlations between the two data sets over the N overlapping samples. We also calculated the statistical power for each test, by calculating the percentage of the time the corresponding $P \leq 0.05$. In this analysis, we assumed that in both data sets, the drug response data has a Gaussian distribution with $N(0,\sigma^2)$ for insensitive and $N(4\sigma,\sigma^2)$ for sensitive cell lines.

ANOVA. The ANOVA was performed on the data set corresponding to the overlapping set of cell lines and using the genomic and tissue annotations and methods described in Garnett *et al.*[9].

Specifically, a vector of length n consisting of AUC (respectively IC_{50}) scores for n cell lines was constructed for each drug. A linear (no interaction terms) ANOVA model was then fitted to these scores with factors including the cell line tissue type and the mutation status of 71 cancer genes, in turn. Significance and effect size (computed by using the Cohen's D) were obtained for each of the gene-drug pairs. This effect size measures the relative difference in the average AUC (respectively IC_{50}) from the wild-type to mutant group compared to the AUC (respectively IC_{50}) pooled standard deviation of the two groups. P values were subsequently corrected for multiple hypothesis testing with the Benjamini–Hochberg method and a threshold of 20% FDR was used to identify significant associations. Subsequently, systematic unpaired Welch's t-tests were performed to identify tissue/drug-response associations for the drugs showing response differences across different tissues, according the ANOVA models.

Elastic net (EN). Since the IC_{50} is not reported in CCLE when it exceeds the tested range of 8 μM, we used the activity area for the regression as in the original CCLE publication. We also used the values considered to be the best in the original GDSC study: the interpolated $\log(IC_{50})$ values. This setting might not be the most comparable to the CCLE study, but it was felt to be more powerful from the standpoint of detecting bona fide associations. In order to compare features between the two studies, we used the same genomic data set (CCLE).

For the GDSC regression we matched the CCLE genomic features cell line names to the GDSC cell line names. From the CCLE genomic features we used 18,900 gene expressions, 1,643 genes probed for mutation (excluding 5′ UTR, introns, 3′ UTR, silent mutations), copy numbers for 446 genes (the Cosmic cancer census genes) and 24 tissues. Elastic net regression was performed as described in Garnett *et al.*[9], using 100 independent runs. For each iteration, the data are cross-validated with a random tenfold partition of the samples. Elastic net performs feature selection: unselected features have zero coefficients. Frequency is the proportion of models out of the 100 runs where the coefficient is non-zero. Note that when we restricted the drug responses to overlapping cell lines/drugs between the two studies, only 12/15 drugs had any features, both in the CCLE regression and the GDSC regression. Contingency tables of sensitive and resistant features in both studies are in Extended Data Fig. 6b. To assess significance of the overlap between the results, χ^2 statistics were computed on the two-by-two contingency tables of sensitive and resistant features (Extended Data Fig. 6a).

Ridge regression. After performing elastic net regression as described above, the genomic features identified were applied to a ridge regression using the same responses as in the elastic net, that is, the IC_{50} values for GDSC (including extrapolated values) and activity area for CCLE. Ridge regression was performed using all features selected by elastic net (Fig. 2b–d and Extended Data Fig. 7). In all plots the axes represent the weights attributed in the ridge regressions that were multiplied by the standard deviation of the features as in Garnett *et al.*[9], and then standardized per drug. The data points are shaded according to their elastic net frequency.

Several measures of consilience of drug-feature associations between the two studies were then computed (Extended Data Fig. 8). In order to gain statistical power, we also performed elastic net regression using all cell lines available in each study separately, and compared the results (Extended Data Fig. 8, second row). Consilience was computed using agreement proportion (that is, the proportion of drug–feature associations with the same sign between the two studies), cosine correlation and Spearman correlation. Cosine correlation conserves the sign of the compared values before scalar product and is therefore a better measure of consilience when only a few features are available for comparison (see for instance nutlin-3 in CCLE panels, Extended Data Fig. 8).

Waterfall method for categorization of cell lines. Our implementation of the waterfall method follows the steps introduced in Barretina *et al.*[8] and described by Haibe-Kains *et al.*[7], with one exception: we only distinguished between sensitive and resistant cell lines. Hence, we removed the intermediate group altogether for simplicity, which allowed for a more straightforward comparison of the two studies. The specific approach we used is outlined as follows.

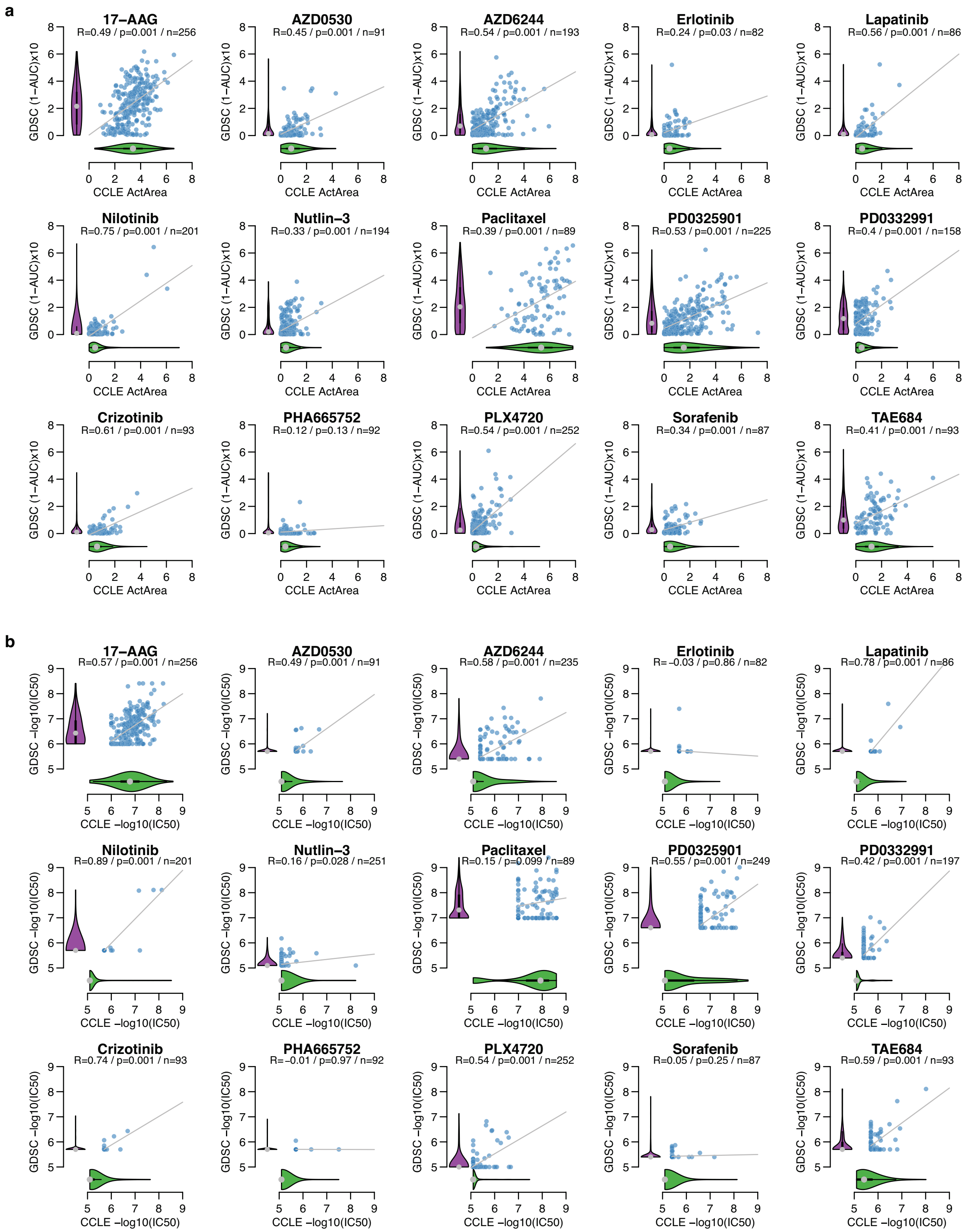
The drug sensitivity measurements were extracted (IC_{50} values for all cell lines measured in the GDSC and CCLE studies). This is a major difference to the Haibe-Kains *et al.* analysis[7], as that analysis only considered the cell-lines in common between the studies when generating response distribution curves. Increasing $\log_{10}(IC_{50})$ values were then sorted to generate a waterfall distribution. If the waterfall distribution is nonlinear (Pearson correlation coefficient to the linear fit ≤ 0.95), the inflection point of the $\log_{10}(IC_{50})$ curve was estimated as the point on the curve with the maximal distance to a line drawn between the start and end points of the distribution. If the waterfall distribution appears linear (Pearson correlation coefficient >0.95), the median $\log_{10}(IC_{50})$ was used instead. Cell lines with $\log_{10}(IC_{50})$ below this inflection point were classified as sensitive, whereas the rest were deemed resistant (see the estimated inflection points in Supplementary Data 2).

Using this approach, we generated drug sensitivity calls for all cell lines within the GDSC and CCLE studies and employed the Cohen's Kappa statistical analysis of agreement.

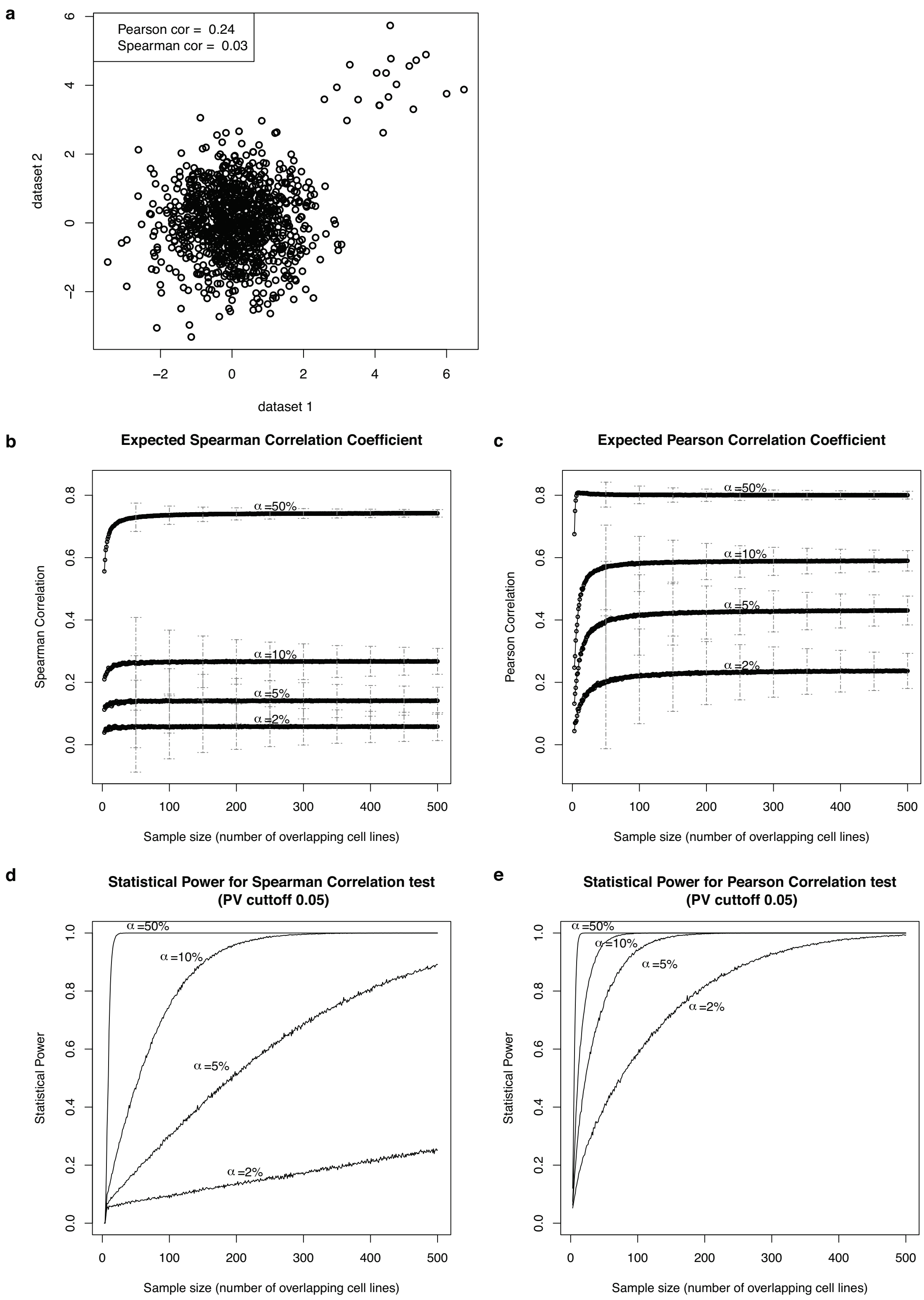
Drug sensitivity analysis scenarios. Using the waterfall method, we estimated inflexion points (which ultimately differentiate between sensitive and resistant cell-lines), on all available/measured cell-lines (for a given drug), in order to have a more complete drug response curve and as a result, better sensitivity agreements (blue bars in Extended Data Fig. 3).

Without the waterfall method, we used a fixed threshold of 1 μM for each drug, in order to distinguish between sensitive and resistant cell-lines. This was much simpler and faster than the previous approach, while generating similar results (green bars in Extended Data Fig. 3).

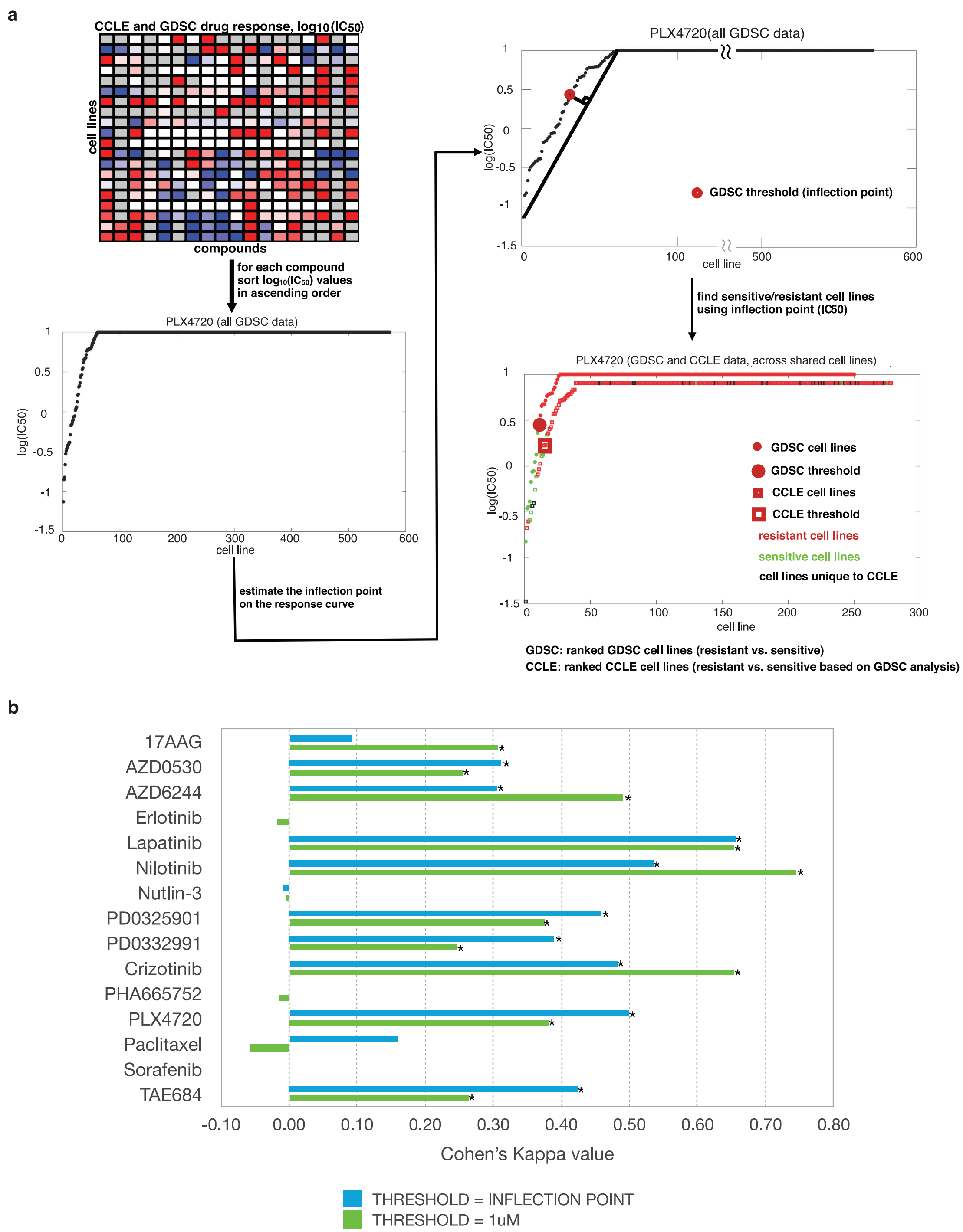
Code availability. Most of the analyses performed in this paper were implemented in R. The R package containing the source codes can be accessed at http://www.broadinstitute.org/ccle/Rpackage.



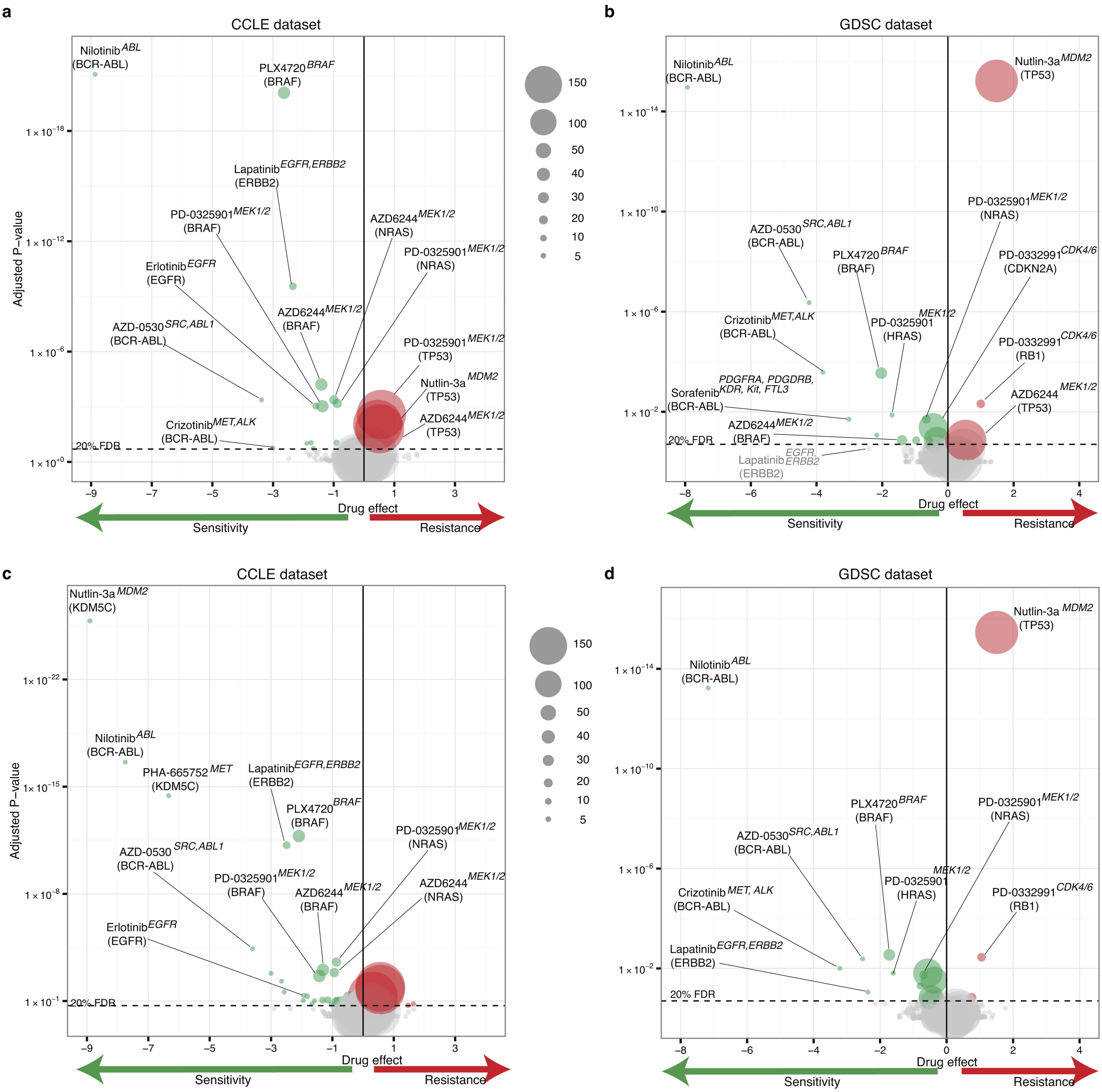
Extended Data Figure 1 | Comparison of pharmacological data from the CCLE and GDSC studies. a, **b**, Scatter plots (blue dots) represent the drug sensitivity measured as the area under the dose–response curve (**a**) and IC_{50} (**b**) in overlapping cell lines between CCLE and GDSC studies. For this analysis, IC_{50} values for insensitive compounds were set to the highest concentration tested in both data sets. The number of overlapping cell lines *n* for each drug is indicated, as well as the Pearson correlation coefficient *R* and *P* value. In this representation, lower values denote insensitive cell lines. The full distribution of sensitivity values for each drug and study is depicted as 'violin plots' (green, CCLE; purple, GDSC) and accounts for all tested cell lines, as opposed to the overlapping set; the grey dot represents the median, thick black line represents the first to third quartile range, and shape of the plot represents the kernel density of the distribution.



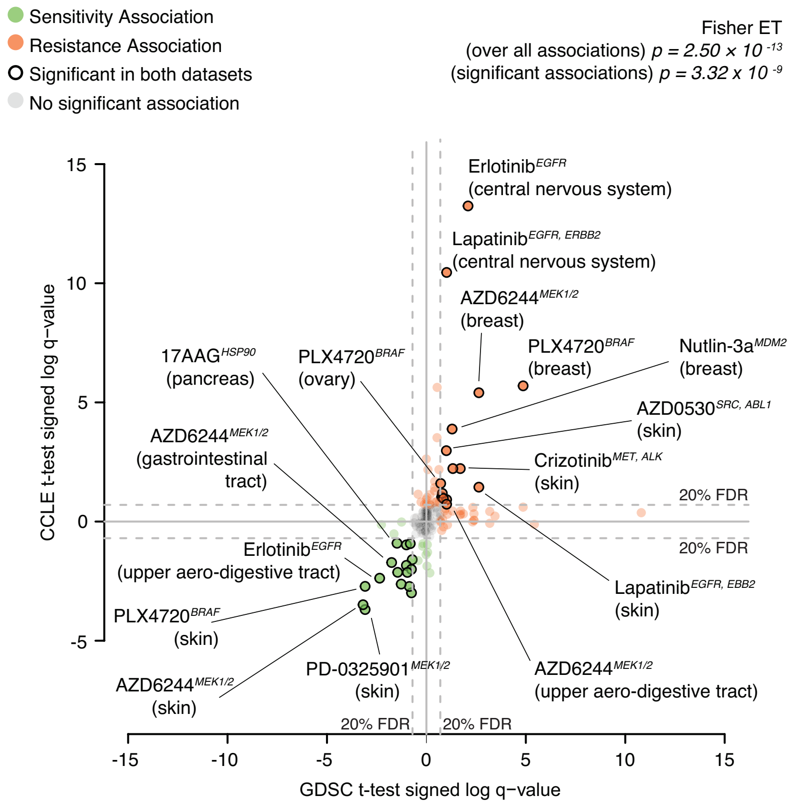
Extended Data Figure 2 | Power analysis of Spearman and Pearson correlation tests. a, Example of a clear signal that appears in only 2% (20 out of 1,000) data points using synthetic data. The Spearman statistic completely fails to detect such a signal which is typical for selective cancer therapeutics. **b**, **c**, Expected Spearman and Pearson correlation coefficients between the two data sets assuming different percentages of drug-sensitive cell lines (α = 2%, 5%, 10% and 50%) and different number of overlapping cell lines. The error bars depict ± one standard deviation. **d**, **e**, Estimated statistical power for Spearman and Pearson correlation tests using a *P* value cutoff of 0.05 for rejecting the null hypothesis. This analysis was done using synthetic data as described in the Methods.



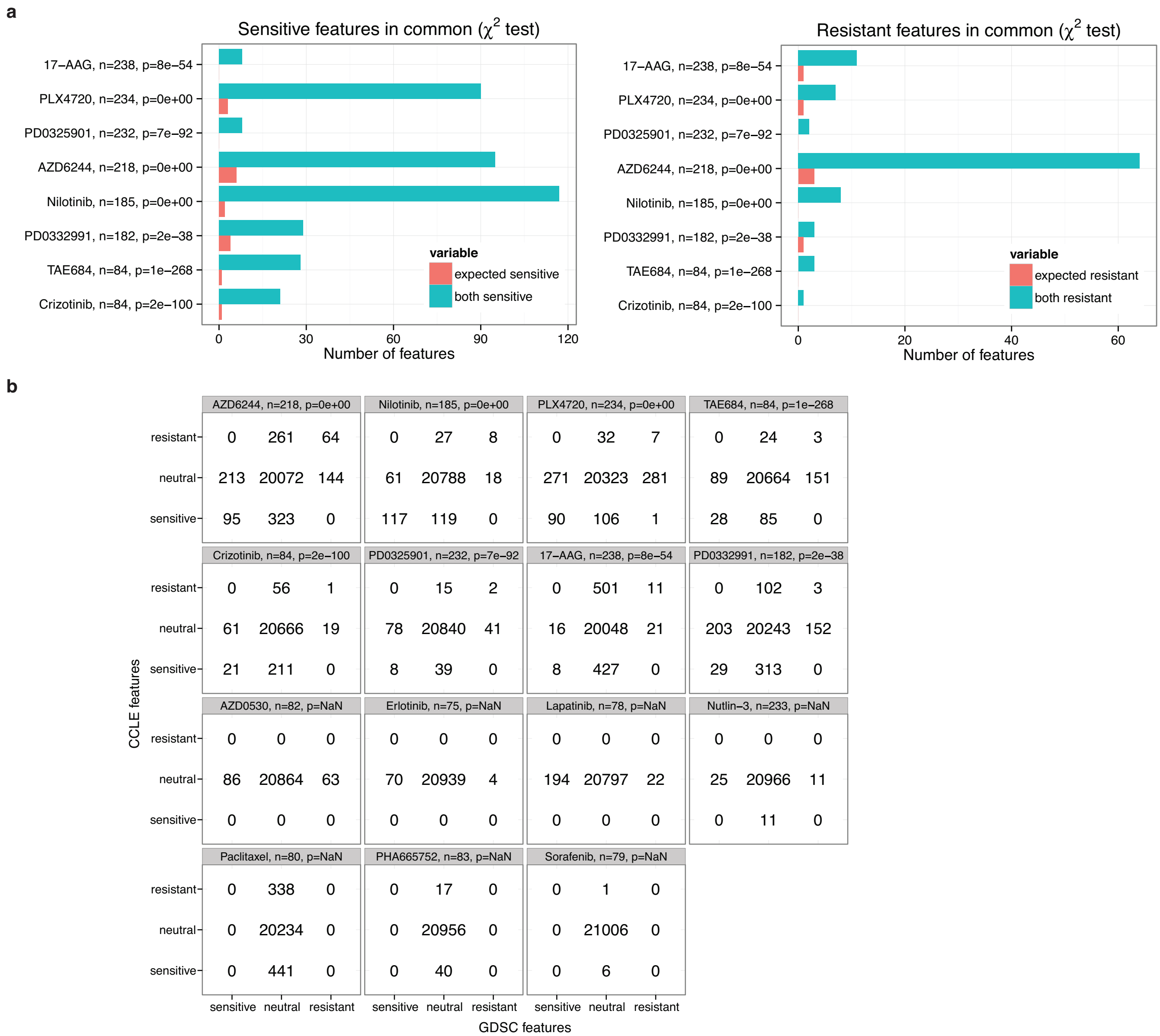
Extended Data Figure 3 | Waterfall analysis for categorization of cell lines. a, Schematic of the waterfall analysis methodology and example of outcome for PLX4720. **b**, Consistency in cell line sensitivity categorization for all drugs. The waterfall method using all data available was used to determine thresholds between 'sensitive' and 'resistant' cell lines (blue). Alternatively a 1 μM threshold was used (green). Asterisks indicate significance of Cohen's Kappa coefficients ($P < 0.05$).



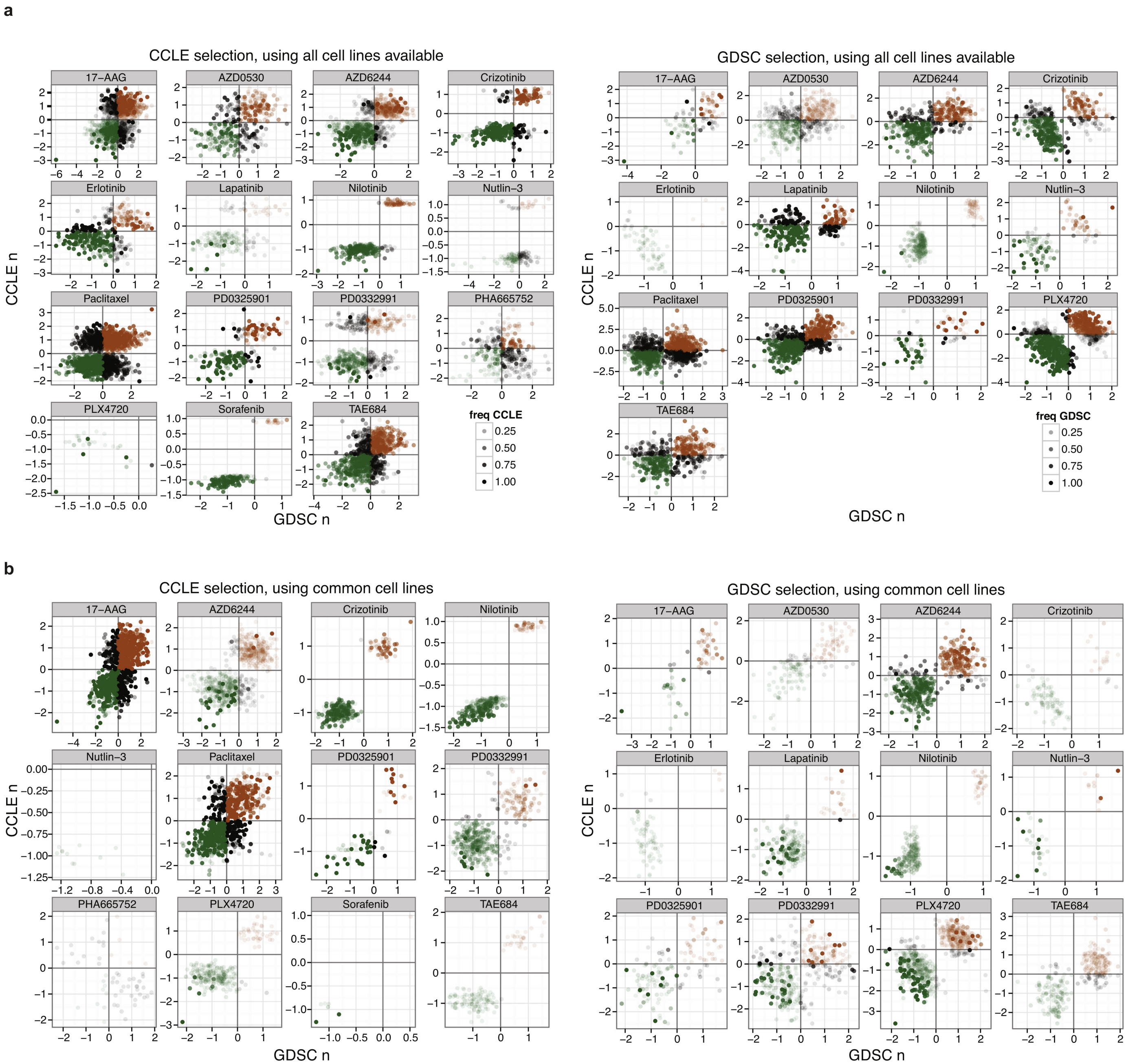
Extended Data Figure 4 | Overlap in ANOVA genomic correlates of drug sensitivity. a–d, Volcano plots showing ANOVA outcomes using drug responses from CCLE (left, **a**, **c**) or GDSC (right, **b**, **d**) data set from overlapping set of cell lines, and mutational status of 71 cancer genes from the GDSC. **a**, **b**, Analyses using AUC values. **c**, **d**, Analyses using IC_{50} values. Points represent drug–gene interactions (with sizes proportional to the number of screened mutant cell lines). Positions on *x* axis indicate effect size magnitudes: negative values (green circle) indicate mutations associated with increase in sensitivity, positive values (red circle) mutations associated with increased resistance. Positions on *y* axis indicate association significances (corrected *P* values) and the horizontal dashed line indicates a significance threshold (FDR 20%). Corresponding drug name, target(s) and cancer gene are reported for a subset of therapeutically relevant interactions.



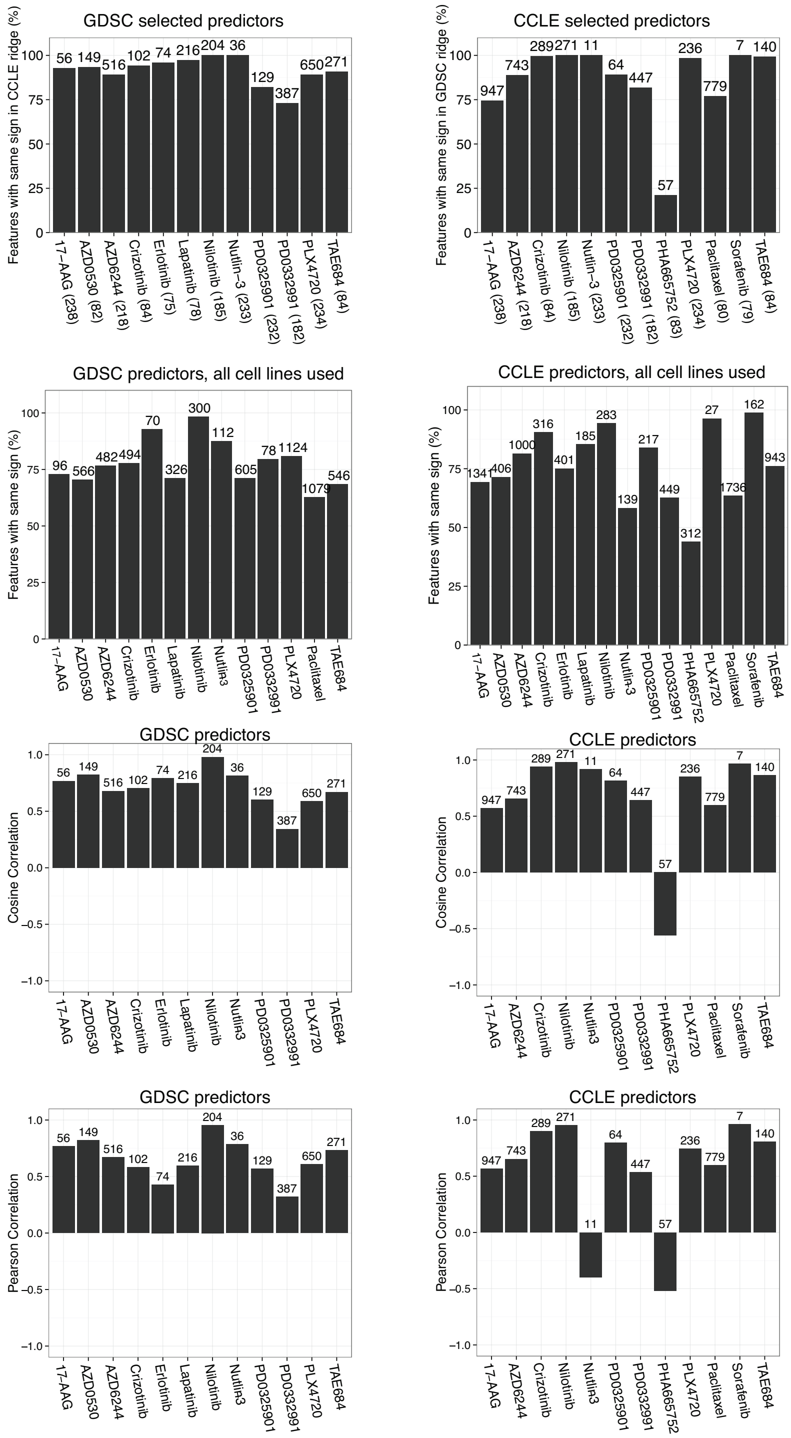
Extended Data Figure 5 | Consistency of drug sensitivity/tissue-of-origin associations between the CCLE and GDSC data sets. Each point is a tested association between drug response and a given cell line's tissue of origin. Positions of the points on the two axes correspond to 'signed log *q*-values' of the corresponding tests for the two data sets, respectively. Point labels indicate drug names and targets (in italics) and tested tissue (among round brackets). The sign indicates the effect of the marker (neg = increased sensitivity and pos = increased resistance) and the magnitude indicates the log *P* value of the corresponding *t*-test, after correcting for multiple hypothesis testing. Fisher's exact test *P* values for independence of columns and rows of the contingency table determined by sign and significance of the associations are also reported (over all the tests and for significant associations only, respectively).



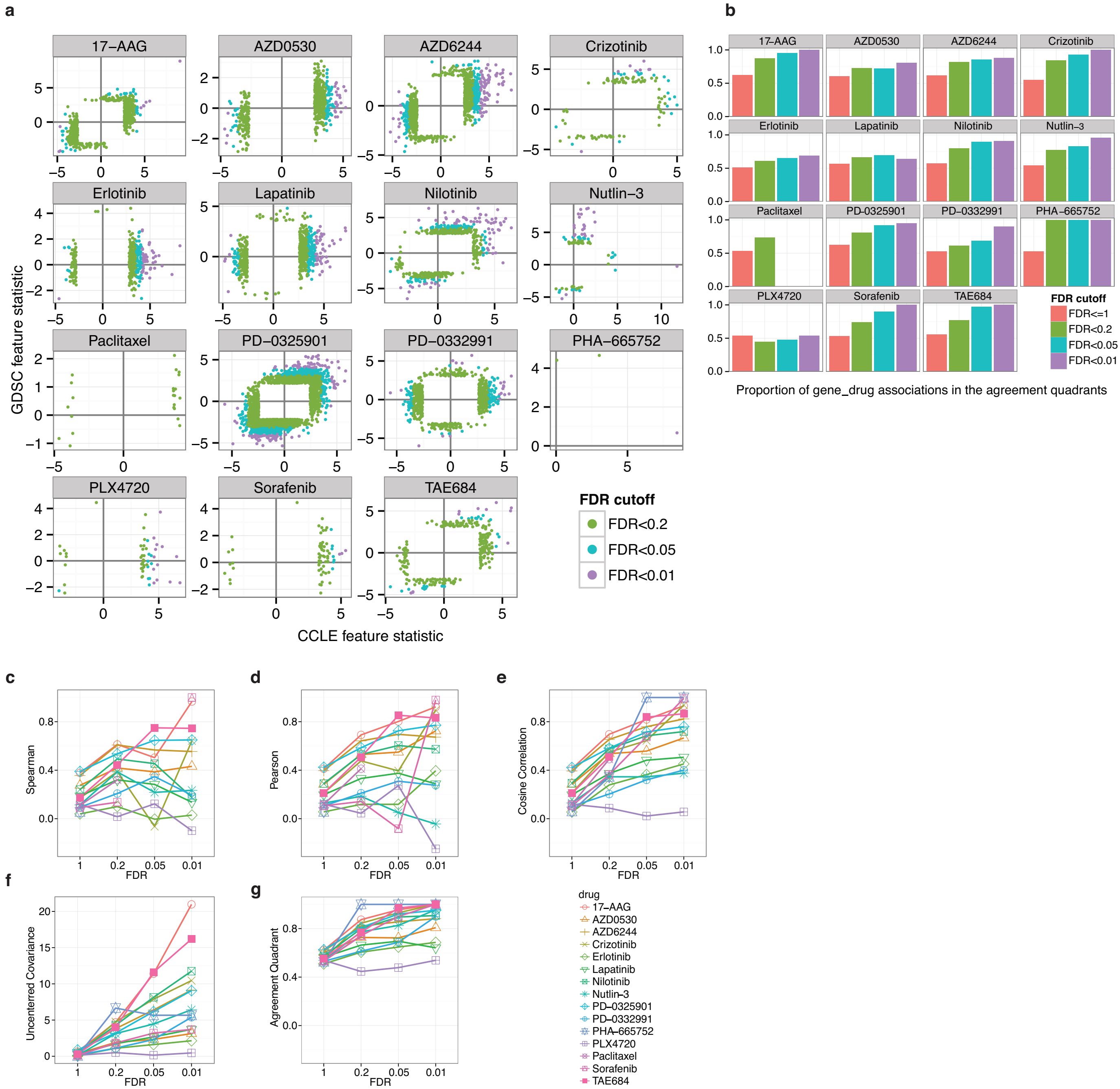
Extended Data Figure 6 | Comparison of genomic features selected by elastic net between the CCLE and GDSC data sets. a, Consistency in predictors of response identified by elastic net regression across 21,013 genome features (copy number variations, messenger RNA expression and sequence variants). Statistical significance of the number of genomic features identified in common (χ^2 test) using the GDSC and CCLE drug sensitivity data sets. Only drugs where features were found in both studies are represented. **b,** Corresponding contingency tables. Out of the 4,957 drug–gene associations with non-zero elastic net weight coefficients, only one divergent result was found (weight coefficient with opposite signs), corresponding to a feature with the lowest possible frequency (non-zero coefficient in 1 out of 100 bootstrap trials in the elastic net analysis).



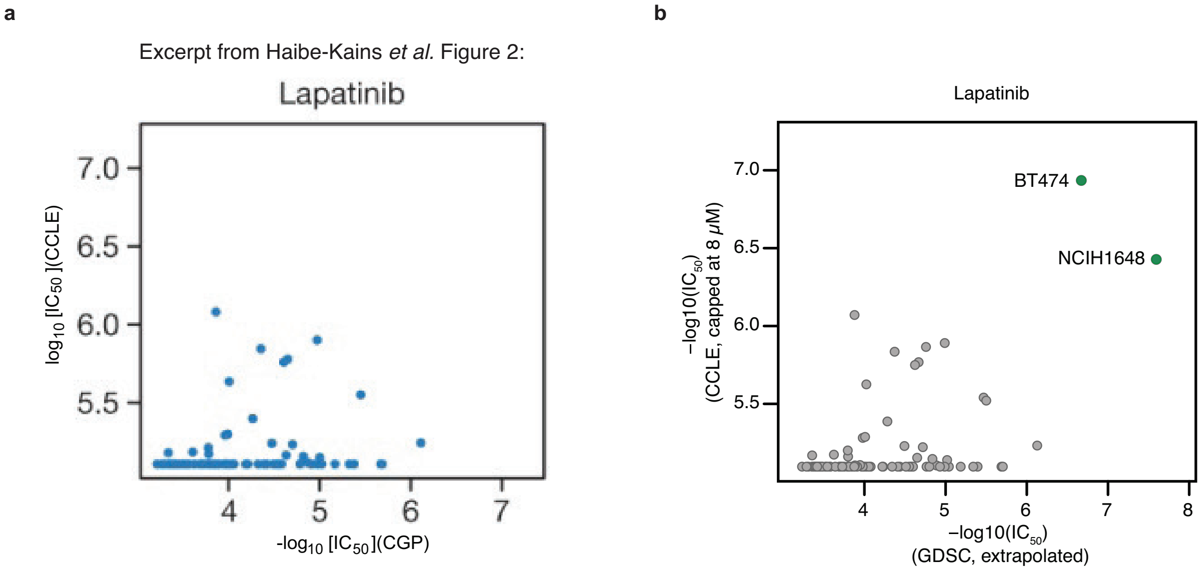
Extended Data Figure 7 | Comparison of genomic feature-drug associations in the CCLE and GDSC data sets. a, **b**, Ridge regression coefficients for all the drugs with successful elastic net regression in the indicated data set are plotted using either overlapping (**a**) or all available (**b**) cell lines. To select cell line features, elastic net was performed using the indicated data set. Then, ridge regression was performed on each data set using the selected features. For plotting, the weights associated with the features were multiplied by the standard deviation of the features as in Garnett *et al.*[9], and then standardized per drug. Colour scale indicates the number of times a feature is selected in 100 independent runs of the elastic net. Green and red colouring indicate features associated with sensitivity or resistance, respectively.



Extended Data Figure 8 | Agreement in genomic predictors of drug response identified by elastic net regression in the GDSC and CCLE studies. Elastic net selection of genomic features was performed on the indicated data set and their effects were computed using a non-selective regression (ridge). Total number of features selected by elastic net is reported above the bars. Number of cell lines used in the regression is in parentheses on the *x* axis. Consistency is reported as the proportion of features with the overall same direction of effect (association with sensitivity or resistance): proportion of features with same sign, using either the cosine correlation that takes into account the sign associated with the features or the Pearson's correlation that does not.



Extended Data Figure 9 | Gene expression correlates of drug response identified previously have better agreement when using more stringent FDR cut-offs. Data from Haibe-Kains *et al.*[7]. **a**, Scatter plots of the IC_{50} based gene-drug association statistic (column "stat" in Haibe-Kains *et al.*[7]; Supplementary Data 2 and 3 and Extended Data Fig. 6) with FDR between 0 and 0.01 (purple), 0.01 and 0.05 (cyan), 0.05 and 0.2 (green). In each panel the two black lines intersect at the origin and define the agreement quadrants (top right and bottom left quadrants). **b**, Proportion of genes in the agreement quadrants (same sign between the two studies). **c**, Additional measures of agreement between the two studies: Agreement measures increase with more stringent FDR cut-off, suggesting that false discovery drives agreement down. Uncentred measures (cosine correlation, uncentred covariance, agreement quadrant proportion) yield better agreement between the studies (see Supplementary Discussion).



Extended Data Figure 10 | Example of significant change in observed correlation by addition of a few sensitive cell lines. For lapatinib sensitivity data, there are 86 overlapping cell lines between the CCLE and GDSC data sets. **a**, Left panel is an excerpt from Haibe-Kains *et al.*[7] figure 2 comparing the sensitivity data of lapatinib for the two data sets. **b**, Right panel shows the two sensitive cell lines (BT-474 and NCI-H1648) that were omitted in the analysis of Haibe-Kains *et al.*[7]. The inclusion of these two cell lines drastically changes the observed Pearson correlation (from 0.25 to 0.53). This is consistent with the simulation results (Extended Data Fig. 2c) that show high variability in the observed Pearson correlation for low sample numbers.

doi:10.1038/nature16144

Thermal biases and vulnerability to warming in the world's marine fauna

Rick D. Stuart-Smith[1], Graham J. Edgar[1], Neville S. Barrett[1], Stuart J. Kininmonth[1,2] & Amanda E. Bates[3]

A critical assumption underlying projections of biodiversity change associated with global warming is that ecological communities comprise balanced mixes of warm-affinity and cool-affinity species which, on average, approximate local environmental temperatures. Nevertheless, here we find that most shallow water marine species occupy broad thermal distributions that are aggregated in either temperate or tropical realms. These distributional trends result in ocean-scale spatial thermal biases, where communities are dominated by species with warmer or cooler affinity than local environmental temperatures. We use community-level thermal deviations from local temperatures as a form of sensitivity to warming, and combine these with projected ocean warming data to predict warming-related loss of species from present-day communities over the next century. Large changes in local species composition appear likely, and proximity to thermal limits, as inferred from present-day species' distributional ranges, outweighs spatial variation in warming rates in contributing to predicted rates of local species loss.

The inherent vulnerability of ecological communities to global warming, and therefore the magnitude of associated biodiversity change, is considered a function of exposure and sensitivity to warming, coupled with species' adaptive capacity[1–3]. Geographic models of future biodiversity change generally accommodate the magnitude, direction and distribution of temperature change[4–8], but have limited ability to account for the sensitivity of communities to change. Our understanding of sensitivity to warming has been largely based on results of comparative studies of species physiological tolerances and other life-history traits, often with extension from the laboratory to the field[9–12]. Extrapolation to whole ecological communities and large geographic scales, does, however, introduce substantial uncertainty, yet these are the scales critical for understanding natural ecosystem functioning[13], on which the well-being of human society depends.

The few studies that have considered community-level sensitivity to warming[3,7,14] have not accounted for geographic patterns in species distributions, inherently assuming that communities comprise balanced mixes of relatively warm-affinity and cool-affinity species, and with no spatial trends or regional consistency in any deviation from this. Regional variation in species composition may be influenced by numerous historical, ecological and phylogenetic factors that could potentially result in thermal bias of communities in relation to local environmental temperatures, with important implications for community-level sensitivity to warming. If, for instance, most species have a warmer affinity than the mean local temperature, then the local community may have little intrinsic sensitivity to negative change with warming. In this case, proxies previously used for inferring sensitivity, such as habitat type or integrity[3], may provide limited predictive insight. Quantifying the direction and magnitude of community thermal bias is therefore an important step in improving our understanding of the sensitivity of ecological communities to structural reorganization with warming, and providing a more direct means to account for sensitivity in predictions of vulnerability.

Thermal biogeography

The community temperature index (CTI) is a measure (a community-weighted mean) of the average thermal affinity of ecological communities, and has recently been used to quantify warming in birds[15,16], butterflies[17] and fishes[18], and global commercial fisheries catches[19]. Here we use the CTI of shallow-water marine fishes and invertebrates to test for thermal bias in the global distribution of marine communities in relation to local environmental temperatures.

We constructed geographic and thermal distributions for 2,695 reef fish and 1,225 mobile macroinvertebrate species using occurrence records from two of the world's most comprehensive databases for shallow-water marine species (Global Biodiversity Information Facility, http://www.gbif.org, and Reef Life Survey[20,21], http://www.reeflifesurvey.com), combined with remotely sensed long-term mean sea surface temperature (SST)[22]. We used the midpoint of the realized thermal distribution as a measure of the central thermal tendency for each species, or thermal affinity. On average, this aligns with the temperature at which species occur at their maximum abundance in the field (see Methods), and is therefore a good proxy for the temperature of a species' maximum ecological success.

We then compiled the first global-scale data set of abundance-weighted CTI values from systematic quantitative sampling, using abundance data for all fish and invertebrate species recorded on standardized visual censuses at 2,447 sites by the Reef Life Survey (RLS) program (see Methods; Extended Data Fig. 1). This approach thus incorporates patterns in species' dominance related to thermal affinity.

A nonlinear global pattern is evident in CTI values, with relatively little change with increasing temperature in tropical and temperate regions, and a rapid increase in subtropical regions creating a distinct step (Fig. 1 and Extended Data Fig. 2a, b). This pattern is consistent between fishes and invertebrates (Pearson correlation $= 0.98$; $n = 2,383$; $P < 0.01$) and is the same when CTI is calculated without weighting by abundance (that is, using presence data; Extended Data Fig. 2c, d). A direct result of this nonlinearity in global CTI is that the majority of locations are characterized by marine communities with either higher or lower CTI than would be expected from local SST (Extended Data Fig. 3). Thermal bias is ubiquitous among these communities, which are typically numerically dominated by species with warmer or cooler affinity than the local environment.

[1]Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania 7001, Australia. [2]Stockholm Resilience Centre, Stockholm University, Kräftriket Stockholm, 2B, SE-106 91, Sweden. [3]Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, Southampton SO14 3ZH, UK.

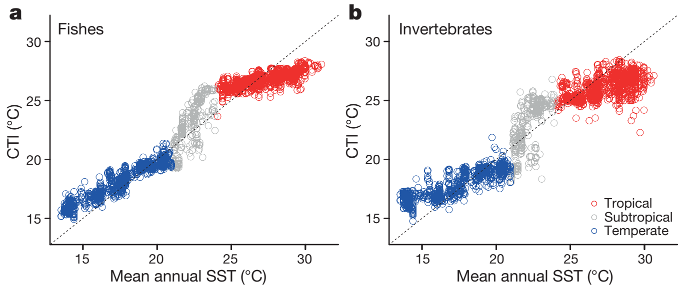


Figure 1 | Global community temperature index values for reef fishes and invertebrates against mean annual sea surface temperature. **a**, **b**, Tropical and temperate communities are separated by subtropical transitions in which communities largely comprise a mixture of temperate and tropical species. A line with a slope of one is plotted for reference. $n = 2{,}175$ and $n = 1{,}901$ sites for fishes and invertebrates, respectively, after exclusion of sites with confidence scores <2.5 (see Methods).

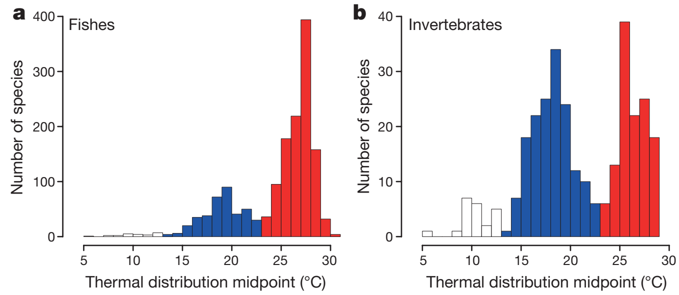


Figure 2 | Frequency distributions of fish and invertebrate species according to their thermal distribution midpoint show modes of temperature affinity or tropical (red), temperate (blue) and subpolar (white) thermal guilds. **a**, **b**, Species for which confidence in thermal midpoints was low are excluded (see Methods).

The proximate cause of large-scale patterns of thermal bias is that marine species distributions do not follow the monotonic latitudinal and temperature gradients observed in species richness[23,24]. Instead, we find that the majority of species studied have ranges centred in either temperate or tropical zones (Extended Data Fig. 4), and consequently show a corresponding multimodal distribution of the thermal affinities (that is, thermal guilds; Fig. 2). This trend is consistent when considered for different ocean basins and biogeographic regions. Additional to the major temperate/tropical dichotomy, the invertebrate data suggest the presence of a third, subpolar thermal guild (Fig. 2b).

Thermal guilds align with the theory that temperature can be considered as an ecological resource in freshwater fishes[25], and can be distinguished within other independent data sets of marine species (see Supplementary Information). The findings of globally coherent thermal guilds is not the result of spatial sampling structure of the data, such as a consequence of relatively few surveys in the subtropics; a latitudinal transect along the well-surveyed north–south trending eastern Australian seaboard clearly distinguishes tropical from temperate faunas along the full cline (Extended Data Fig. 5). There are several potential, non-mutually exclusive mechanisms that may explain these findings: (1) fewer shallow-water species may have ranges centred in subtropical ocean climates as a result of less continental shelf area at subtropical latitudes globally[26]; (2) historical biogeographic processes could be implied for the Australian fauna, through mixing of tropical Pacific/southeast Asian and temperate Australian faunas as the Australian continental plate drifted north, with species conserving thermal preferences (that is, phylogenetic inertia[27]); (3) tropical centres of speciation and subsequent colonization of temperate regions through 'bridge species' may have occurred (the 'out of the tropics' hypothesis[26]), and is supported by the distributions of thermal affinities of species in large families of fishes that span temperate and tropical zones (Extended Data Fig. 6); (4) there could be adaptive advantages associated with specialization for either warm or cool temperature ranges, with trade-offs in metabolic processes reducing widespread adaptation to intermediate temperatures.

Regardless of the ultimate drivers, the existence of consistent thermal guilds and associated global-scale patterns of thermal bias has implications for whether the net community response to warming is more likely to be positive or negative (in terms of abundance changes). It also raises the possibility that communities in some locations may be more vulnerable to losing species than in other locations, simply on the basis of the direction and magnitude of the bias in the thermal distributions of the species present.

Vulnerability of marine communities to warming

Most previous biodiversity vulnerability analyses have focused on species, and their ability to change their geographic distribution or adapt to avoid global extinction[10,28]. Here we quantitatively assess the vulnerability of whole communities—groups of species that are currently recorded as co-occurring and interacting at an ecologically relevant scale. A local ecological community is considered vulnerable if it is likely to lose many of its constituent species. This may not translate to reductions in overall species richness (although see below), but does reflect a relative vulnerability to change in community structure and ecosystem functioning, and contrasts with desirable management goals of resilience or stability in the face of warming[29].

Over decadal scales, positive thermal bias of the magnitude observed for some locations in this study (for example, where the mean thermal affinity of the community is 3 °C greater than local mean SST) is much greater than predicted ocean warming rates of <0.4 °C per decade, and may translate to low probabilities of species loss as a result of warming, or relatively low community sensitivity to negative change. Most species in such locations are also found in other warmer locations, and so are unlikely to be negatively affected by warming. However, the likelihood of local loss of species on the basis of increasing temperature will be more dependent on how close each of the species is and becomes, at that location, to the maximum of its thermal distribution, rather than from the midpoint (as used to define thermal bias in our thermal biogeographic analysis). To account for this, we recalculated CTI using the 95th percentile of species' thermal distributions as a measure of contemporary realized upper thermal limits (CTI_{max}). Realized upper limits will be lower than fundamental limits based on physiological tolerances, but arguably better reflect real-world limits, where species not only need to survive physiologically, but also persist in a competitive and predatory environment.

For calculation of CTI_{max} to estimate species loss with warming, we used presence rather than abundance data and combined RLS survey data for fishes and invertebrates, thereby covering the majority of macroscopic mobile fauna (>2.5 cm) on rocky and coral reefs at sites investigated. We re-calculated thermal bias ($TBias_{max}$) as the difference between CTI_{max} and mean summer temperatures (mean SST from the 8 warmest weeks annually from 2008–2014 (ref. 30)). This can be considered a form of 'distribution safety margin'[27], and shows a similar global pattern to that shown in our thermal biogeographic analysis (Extended Data Fig. 7), with CTI_{max} and CTI very closely related (Pearson correlation = 0.96; $n = 2{,}089$; $P < 0.01$).

CTI_{max} also shows a stepped relationship with summer SST (Extended Data Fig. 8), reflecting some consistencies among species' realized upper thermal limits within tropical and temperate regions at the global scale. For example, CTI_{max} remains between 22 °C and 24 °C across most sites with summer temperatures ranging from 14 °C to 24 °C, implying that the average species is living closer to their warmest distributional margin at locations with summer temperatures around approximately 24 °C than at locations which experience summer temperatures around approximately 14 °C. $TBias_{max}$ is consequently more positive for the latter, although sites dominated by species in the tropical thermal guild (as identified in Figs 1 and 2) that experience summer temperatures around approximately 24 °C (that is, on the

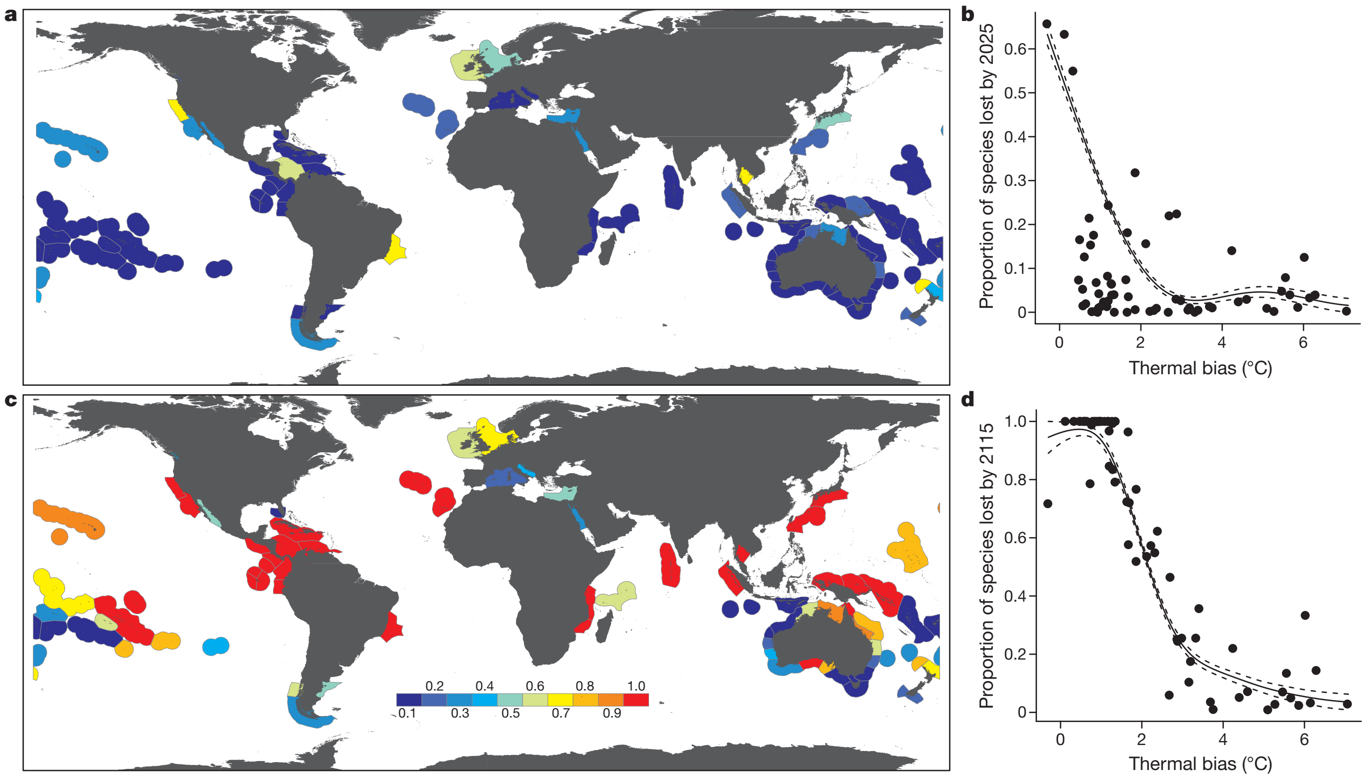


Figure 3 | Vulnerability of marine communities to warming-related local species loss. a–d, Proportion of fish and invertebrate species in present-day communities likely to exceed their upper realized thermal limit by 2025 (**a**) and 2115 (**c**) based on regional IPCC warming rates (RCP8.5 scenario), and in relation to the magnitude of community thermal bias (measured as $TBias_{max}$; **b**, **d**). Fitted curves (solid black line) and 95% confidence intervals (dotted black lines) are from GAMM models (Extended Data Table 2). Sites with confidence scores <2.5 were excluded from most ecoregion[43] means (see Extended Data Table 1 for sample sizes and details of exclusions).

upper line in Extended Data Fig. 8) also have high $TBias_{max}$ and inferred low sensitivity.

Although $TBias_{max}$ can be considered a form of community-level sensitivity, it does not account for warming rates, another important component of vulnerability[1,2]. To explicitly account for spatial patterns in warming rates and provide quantitative vulnerability predictions for marine communities, we further calculated the proportion of species in the community that would exceed the upper limit of their realized temperature distribution in 10 and 100 years from the present. These are based on each species' contemporary upper thermal limits, recent summer temperatures, and the rate of warming expected at each site (based on ensemble averages from all climate models included in the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) for sea surface temperature anomaly under the RCP8.5 scenario predicted for 2050–2099; http://www.esrl.noaa.gov/psd/ipcc/).

A total of 6 (out of 75) ecoregions included in the analysis were identified in which the mean summer sea temperature is expected to exceed the upper thermal limit of more than 50% of the recorded species by 2025 (Fig. 3a, b). Confidence scores for CTI_{max} values are low for a number of sites in three of these ecoregions on the basis of less comprehensive sampling of species thermal distributions (see Methods and Extended Data Table 1), but were high for sites in the Gulf of Thailand, southwestern Caribbean and Three Kings-North Cape (New Zealand). Longer-term predictions are more extreme, with 100% of the present-day community composition apparently likely to exceed upper thermal limits in approximately one-third of surveyed ecoregions by 2115 (Fig. 3c, d). These are distributed in all ocean basins across the tropics, but also in some temperate areas such as the Great Australian Bight.

Locations of greatest predicted species loss do not closely align to locations of greatest warming, but instead correspond closely to the magnitude of thermal bias (measured as $TBias_{max}$; Fig. 3b, d; GAMM results in Extended Data Table 2). This result is robust to the warming data used (see Supplementary Information), and shows that sensitivity associated with community thermal bias is an important component of vulnerability. Our results further indicate that exposure, and variability in warming rate predictions, may be considerably less important than previously suggested[1] when it comes to local loss of marine species over the next century. Predicted species loss at locations with lower thermal bias is considerably greater than at locations with higher thermal bias, despite some of the world's most rapidly warming regions occurring within the latter. The western Mediterranean, for example, is predicted to warm by 0.24–0.29 °C per decade (depending on predictions used), but typical marine communities there consist of species with contemporary upper limits well above local summer SST (mean $TBias_{max} = 6.3\,°C \pm 1.1$ s.d.).

Our predictions do not account for local influx of warmer affinity species, and do not comprise the only form of community-level vulnerability to warming. Rather, they describe effects of an additional component of ecological vulnerability. Species influx and warming-associated changes in species abundances will also contribute to local ecological change and are already occurring in the most rapidly warming areas that are well-connected to rich tropical faunas, such as south-eastern Australia[12]. An influx of warm-affinity species may replace lost species or lead to accumulating richness in some regions, and probably have dramatic impacts on ecological processes[6,31]. Local species loss through extinction or range contraction will represent the main form of community change probable for low-latitude regions for which no pool of warmer affinity species exists[11,32], however, and so our predictions probably cover the major changes in composition expected in these regions.

A key assumption for our vulnerability analysis is that local extinction becomes more probable when a site becomes warmer than the typical maximum temperature at which a species has previously been

observed. This assumption relies on the interactive mechanisms which presently set boundaries on species' ranges remaining consistent, such as thermally driven performance reduction[33,34] and increased susceptibility to competition and predation[18,35]. This is unlikely to be true for all species, especially narrow-range endemics which are probably limited in distribution by factors other than temperature[12]. Regardless, we consider this generalization reasonable given the well-connected nature of the marine environment, typically with large geographic ranges[36], and often closely matching fundamental (assessed in laboratory experiments) and realized (field-derived from distribution data) thermal niches[37], as well as implications associated with lower concentrations of dissolved oxygen in the marine environment with increasing temperature[38].

Our vulnerability predictions also do not account for ecological change resulting from extreme events, which will change biodiversity in spatially variable and largely unpredictable ways. This is particularly true for indirect effects of extreme events, such as through habitat change, which place critical pressures on biodiversity[39], and represent an important direction for future research.

Additional caveats associated with assessing vulnerability in terms of local loss of species from present-day communities include: (1) the upper thermal limits for many tropical marine species could exceed contemporary ocean temperature maxima, and (2) adjustment and thermal adaptation could reduce species loss from that predicted. The former does not affect results for temperate regions, but could lead to lower vulnerability than predicted for tropical regions, despite results of laboratory experiments that have applied greater temperatures than contemporary SST, suggesting that maximum thermal tolerance levels are more constrained for tropical than temperate species[11,27,40]. Because of these caveats, we emphasize that absolute values presented in Fig. 3 should be considered as a 'worst case scenario' and interpreted with caution. Nevertheless, relative differences in the magnitude of predicted change between regions and times should be robust, other than perhaps overestimation of site-scale species loss at the lowest latitudes relative to cooler climes. Most importantly, the strength of empirical trends indicates that thermal bias is a fundamental element affecting global variability in future biodiversity change.

Tracking and managing warming impacts on biodiversity

In contrast to prior global studies of potential biodiversity losses associated with climate change, which typically consider loss of species from their full distribution or use regional species lists inferred from range maps, our study focused on probabilities of local-scale losses from assemblages of interacting species. These will be much more pervasive than cases of global extinction, and have important consequences with respect to the way ecosystems currently function. We identify a substantial pressure of warming through the future, with an alarmingly large proportion of species predicted to exceed current realized thermal limits based on current distribution patterns.

Our results imply that locations at which the average summer SST is presently approximately 24 °C are most vulnerable to community change in general. This temperature corresponds to the upper realized thermal limit of many temperate species, and consequently a ceiling on CTI_{max} for most temperate communities. For locations with connections to tropical faunas, it is also where the influx from the large pool of tropical species is going to be greatest. By contrast, the warmest tropical locations are likely to suffer from local loss of species with little replacement, a result consistent among other studies relating biodiversity change to global variation in predicted ocean climate velocity[4,6].

Management options for decreasing local marine species losses resulting from warming are limited; nevertheless, reducing the effects of other threats such as pollution, invasive species, and excessive extraction of living resources, will probably provide the best opportunities for prolonging persistence of species at the warm end of their range. Although some local losses of species appears inevitable, management can bolster community resilience to ocean warming through strategies to reduce influx of warm-affinity species at those regions where accumulation is predicted. Actions to support more intact naturally functioning communities are recommended, including implementation of marine protected areas (MPAs) and more conservative fisheries management. Recent evidence from an effective temperate MPA suggests that local predators hinder poleward progression of warm-affinity species[18], and invasion theory more generally predicts intact and diverse natural communities possess greater resistance to invasive species than degraded communities[41].

Abundance-weighted CTI, as used in our thermal biogeographic analysis, offers an important tool for measuring the success of such management actions, as it integrates signals from local species gains and losses, and also abundance shifts related to temperature. The CTI provides a powerful metric for tracking long-term biodiversity change in relation to warming over larger scales[15], and for informing the wider public of the magnitude of warming impacts on biodiversity. It can thus fill a critical gap in the indicator suite used for assessing progress towards international targets agreed under the Convention on Biological Diversity (CBD). However, we must consider for such application that the magnitude of CTI change will be nonlinear across latitude, with reduced scope for change in tropical regions. The CTI offers an important opportunity to extend emphasis from charts or maps of pressures, such as atmospheric CO_2 concentrations and ocean heat content[42], towards measures of biodiversity change, thereby providing a better understanding of real-life consequences of ocean warming for effective long-term change in policy and human behaviour.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 March 2014; accepted 13 October 2015.
Published online 11 November 2015.

1. Williams, S. E., Shoo, L. P., Isaac, J. L., Hoffmann, A. A. & Langham, G. Towards an integrated framework for assessing the vulnerability of species to climate change. *PLoS Biol.* **6,** e325 (2008).
2. Dawson, T. P., Jackson, S. T., House, J. I., Prentice, I. C. & Mace, G. M. Beyond predictions: biodiversity conservation in a changing climate. *Science* **332,** 53–58 (2011).
3. Watson, J. E. M., Iwamura, T. & Butt, N. Mapping vulnerability and conservation adaptation strategies under climate change. *Nature Clim. Change* **3,** 989–994 (2013).
4. Burrows, M. T. *et al.* Geographical limits to species-range shifts are suggested by climate velocity. *Nature* **507,** 492–495 (2014).
5. Burrows, M. T. *et al.* The pace of shifting climate in marine and terrestrial ecosystems. *Science* **334,** 652–655 (2011).
6. García Molinos, J. *et al.* Climate velocity and the future global redistribution of marine biodiversity. *Nature Clim. Change* http://dx.doi.org/10.1038/nclimate2769 (2015).
7. Beaugrand, G., Edwards, M., Raybaud, V., Goberville, E. & Kirby, R. R. Future vulnerability of marine biodiversity compared with contemporary and past changes. *Nature Clim. Change* **5,** 695–701 (2015).
8. Lima, F. P. & Wethey, D. S. Three decades of high-resolution coastal sea surface temperatures reveal more than warming. *Nature Commun.* **3,** 704 (2012).
9. Foden, W. B. *et al.* Identifying the world's most climate change vulnerable species: a systematic trait-based assessment of all birds, amphibians and corals. *PLoS ONE* **8,** e65427 (2013).
10. Pacifici, M. *et al.* Assessing species vulnerability to climate change. Nature *Clim. Change* **5,** 215–224 (2015).
11. Sunday, J. M., Bates, A. E. & Dulvy, N. K. Global analysis of thermal tolerance and latitude in ectotherms. *Proc. R. Soc. Lond. B* **278,** 1823–1830 (2011).
12. Sunday, J. M. *et al.* Species traits and climate velocity explain geographic range shifts in an ocean-warming hotspot. *Ecol. Lett.* **18,** 944–953 (2015).
13. Kordas, R. L., Harley, C. D. G. & O'Connor, M. I. Community ecology in a warming world: the influence of temperature on interspecific interactions in marine systems. *J. Exp. Mar. Biol. Ecol.* **400,** 218–226 (2011).
14. Okey, T. A., Agbayani, S. & Alidina, H. M. Mapping ecological vulnerability to recent climate change in Canada's Pacific marine ecosystems. *Ocean Coast. Manage.* **106,** 35–48 (2015).
15. Devictor, V., Julliard, R., Couvet, D. & Jiguet, F. Birds are tracking climate warming, but not fast enough. *Proc. R. Soc. Lond. B* **275,** 2743–2748 (2008).
16. Devictor, V. *et al.* Differences in the climatic debts of birds and butterflies at a continental scale. *Nature Clim. Change* **2,** 121–124 (2012).
17. Zografou, K. *et al.* Signals of climate change in butterfly communities in a Mediterranean protected area. *PLoS ONE* **9,** e87245 (2014).

18. Bates, A. E. *et al.* Resilience and signatures of tropicalization in protected reef fish communities. *Nature Climate Change* **4,** 62–67 (2013).
19. Cheung, W. W. L., Watson, R. & Pauly, D. Signature of ocean warming in global fisheries catch. *Nature* **497,** 365–368 (2013).
20. Edgar, G. J. & Stuart-Smith, R. D. Systematic global assessment of reef fish communities by the Reef Life Survey program. *Scientific Data* **1,** 140007 (2014).
21. Edgar, G. J. & Stuart-Smith, R. D. Ecological effects of marine protected areas on rocky reef communities: a continental-scale analysis. *Mar. Ecol. Prog. Ser.* **388,** 51–62 (2009).
22. Tyberghein, L. *et al.* Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Glob. Ecol. Biogeogr*. **21,** 272–281 (2012).
23. Tittensor, D. P. *et al.* Global patterns and predictors of marine biodiversity across taxa. *Nature* **466,** 1098–1101 (2010).
24. Stuart-Smith, R. D. *et al.* Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature* **501,** 539–542 (2013).
25. Magnuson, J. J., Crowder, L. B. & Medvick, P. A. Temperature as an ecological resource. *Am. Zool.* **19,** 331–343 (1979).
26. Jablonski, D. *et al.* Out of the tropics, but how? Fossils, bridge species, and thermal ranges in the dynamics of the marine latitudinal diversity gradient. *Proc. Natl Acad. Sci. USA* **110,** 10487–10494 (2013).
27. Kellermann, V. *et al.* Upper thermal limits of *Drosophila* are linked to species distributions and strongly constrained phylogenetically. *Proc. Natl Acad. Sci. USA* **109,** 16228–16233 (2012).
28. Thomas, C. D. *et al.* Extinction risk from climate change. *Nature* **427,** 145–148 (2004).
29. Mumby, P. J., Chollett, I., Bozec, Y.-M. & Wolff, N. H. Ecological resilience, robustness and vulnerability: how do these concepts benefit ecosystem management? *Current Opinion in Environmental Sustainability* **7,** 22–27 (2014).
30. Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C. & Wang, W. An improved in situ and satellite SST analysis for climate. *J. Clim.* **15,** 1609–1625 (2002).
31. Hiddink, J. G. & Ter Hofstede, R. Climate induced increases in species richness of marine fishes. *Glob. Change Biol.* **14,** 453–460 (2008).
32. Nguyen, K. D. T. *et al.* Upper temperature limits of tropical marine ectotherms: global warming implications. *PLoS ONE* **6,** e29340 (2011).
33. Pörtner, H. O. Climate change and temperature-dependent biogeography: oxygen limitation of thermal tolerance in animals. *Naturwissenschaften* **88,** 137–146 (2001).
34. Pörtner, H. O. & Knust, R. Climate change affects marine fishes through the oxygen limitation of thermal tolerance. *Science* **315,** 95–97 (2007).
35. Figueira, W. F., Biro, P., Booth, D. J. & Valenzuela, V. C. Performance of tropical fish recruiting to temperate habitats: role of ambient temperature and implications of climate change. *Mar. Ecol. Prog. Ser.* **384,** 231–239 (2009).
36. Brown, J. H., Stevens, G. C. & Kaufman, D. M. The geographic range: size, shape, boundaries, and internal structure. *Annu. Rev. Ecol. Syst.* **27,** 597–623 (1996).
37. Sunday, J. M., Bates, A. E. & Dulvy, N. K. Thermal tolerance and the global redistribution of animals. *Nature Clim. Change* **2,** 686–690 (2012).
38. Deutsch, C., Ferrel, A., Seibel, B., Pörtner, H. O. & Huey, R. B. Climate change tightens a metabolic constraint on marine habitats. *Science* **348,** 1132–1135 (2015).
39. Graham, N. A. J. *et al.* Dynamic fragility of oceanic coral reef ecosystems. *Proc. Natl Acad. Sci. USA* **103,** 8425–8429 (2006).
40. Araújo, M. B. *et al.* Heat freezes niche evolution. *Ecol. Lett.* **16,** 1206–1219 (2013).
41. Tilman, D. Community invasibility, recruitment limitation, and grassland biodiversity. *Ecology* **78,** 81–92 (1997).
42. Victor, D. G. & Kennel, C. F. Climate policy: ditch the 2°C warming goal. *Nature* **514,** 30–31 (2014).
43. Spalding, M. D. *et al.* Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *Bioscience* **57,** 573–583 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the many Reef Life Survey (RLS) divers who participated in data collection and provide ongoing expertise and commitment to the program, University of Tasmania staff including J. Berkhout, A. Cooper, M. Davey, J. Hulls, E. Oh, J. Stuart-Smith and R. Thomson. Development of RLS was supported by the former Commonwealth Environment Research Facilities Program, and analyses were supported by the Australian Research Council, Institute for Marine and Antarctic Studies, and the Marine Biodiversity Hub, a collaborative partnership supported through the Australian Government's National Environmental Science Programme. Additional funding and support for field surveys was provided by grants from the Ian Potter Foundation, CoastWest, National Geographic Society, Conservation International, Wildlife Conservation Society Indonesia, The Winston Churchill Memorial Trust, Australian-American Fulbright Commission, and ASSEMBLE Marine.

Author Contributions R.D.S.-S., A.E.B. and G.J.E. conceived the idea, G.J.E., R.D.S.-S. and many others collected the data. R.D.S.-S. drafted the paper, with substantial input from A.E.B., G.J.E., N.S.B. and S.J.K. S.J.K. prepared the maps, A.E.B. and R.D.S.-S. analysed the data and prepared figures.

Author Information A 'live' (periodically updated) database containing the Reef Life Survey ecological data used in this study is accessible online through http://www.reeflifesurvey.com. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.D.S.-S. (rstuarts@utas.edu.au).

METHODS

Reef fish and invertebrate data. Standardised quantitative censuses of reef fishes and echinoderms (holothurians, echinoids, asteroids, crinoids), molluscs (gastropods, cephalopods), and crustaceans (decapods) were undertaken by trained recreational SCUBA divers along 7,040 transects at 2,447 sites worldwide through the Reef Life Survey (RLS) program. Full details of fish census methods are provided in refs 20, 21, and an online methods manual (http://www.reeflifesurvey.com) describes all data collection methods, including for invertebrates. Data quality and training of divers are detailed in ref. 20 and supplementary material in ref. 24. Data used in this study are densities of all species recorded per 500 m^2 transect area for fishes (2 × 250 m^2 blocks), and per 100 m^2 for invertebrates (2 × 50 m^2 blocks). Four per cent of all records were not identified to species level (mostly invertebrates) and were omitted from analyses for this study.

Data from fish and invertebrate surveys were analysed separately for thermal biogeography analyses, but combined for the vulnerability predictions shown in Fig. 3. Although collected on the same transect lines, these survey components cover different areal extents, and so were combined to represent densities per 50 m^2 (block size for invertebrate surveys). Raw invertebrate data were therefore used, but one in five individual fishes were randomly subsampled from those surveyed in each 250 m^2 block to provide equivalent densities and richness of fishes per 50 m^2.

Characterization of species' thermal distributions. A realized thermal distribution was constructed for all species recorded on RLS transects, based on occurrences rather than species distribution models. All individual records within the RLS database were combined with all records of these species in the Global Biodiversity Information Facility (GBIF: http://www.gbif.org/), after applying filters to limit records to depths shallower than 26 m and time of collection since 2004. This resulted in a data set of 399,927 geo-referenced occurrences of 3,920 species.

Remotely sensed local SST data were then matched to each occurrence location. Long-term mean annual SST values from 2002–2009 from the Bio-ORACLE data set[22] were used to provide a time-integrated picture of temperatures species were typically associated with for the thermal biogeographic analysis. The fifth and 95th percentiles of the temperature distribution occupied by each species were then calculated, and the midpoint between these used as a measure of central tendency of their realized thermal distribution. Midpoints were considered a reasonable proxy for the temperature associated with species' maximum ecological success, confirmed by a close alignment of midpoints with the temperatures at which species occurred in maximum abundance in the global RLS data set (slope of midpoint versus temperature of sites at which species were at maximum abundance = 1.003, Pearson correlation = 0.93, $P < 0.001$). Thus, although interspecific variation is expected, deviation in temperatures either side of the midpoint results in reduced abundance for the average species.

We also calculated and explored other metrics from the thermal range, including the median and mode, but these were more sensitive to the distribution and intensity of sampling effort across the temperature range of species, and therefore less robust than the midpoints. Fifth and 95th percentiles were deliberately chosen as endpoints rather than the maximum and minimum because marine species range boundaries are not static, with dynamic tails in distributions[44]. Sightings of individual vagrants are common, sometimes at large distances from the nearest viable populations. Furthermore, any misidentification errors would have greatest influence if at the edge of species ranges.

Community temperature index calculation and thermal bias. CTI was calculated separately for fishes and invertebrates for each transect in the RLS database as the average of thermal midpoint values for each species recorded, weighted by their $\log(x+1)$ abundance. Multiple transects were usually surveyed at each site (2.8 transects global mean across sites used in this study). CTI values were averaged across these to create a site-level mean that was used for analyses. In some cases this averaged out seasonal effects, where sites were surveyed across multiple seasons.

Thermal bias was calculated as the difference between the CTI and mean annual SST at each site. Mean thermal bias values across sites surveyed in each ecoregion are shown in Extended Data Fig. 3, with sample sizes for ecoregions shown in Extended Data Table 1.

Confidence scores. The number of occurrence records for each species ranged from a single record (numerous species) to 1,009 (the Indo-Pacific cleaner wrasse, *Labroides dimidiatus*), with an overall mean of 36 records (47 for fishes, 16 for invertebrates). In order to consider how variation in the comprehensiveness of data on the thermal distribution for each species affected the calculation of CTI and provide an objective measure of confidence in site-level CTI values, we used a semiquantitative confidence scoring system. A confidence value ranging from one (very little confidence) to three (high confidence) was allocated to each species through a four-step process:

(1) The number of records (sites) for each species was used as a first pass for classification, with species observed at 30 or more sites given a value of three, 10–29 sites a value of two, and less than 10 sites, a value of one.

(2) The thermal range for each species (the difference between 95th and fifth percentiles) was used in a second pass for all species that were initially given a value of two. For this, those species with a thermal range of less than 3 °C were reduced to a value of one, as it is possible these species have not been surveyed across their full potential thermal range.

(3) Species with a value of three and a thermal range of less than 1 °C were reduced to a two, given these likely represent well-sampled, but range-restricted species, and their potential thermal range is likely greater than their realized range (which is probably limited by other factors such as dispersal or historical biogeography).

(4) The frequency of occurrences across temperatures was also plotted separately for each species. Frequency histograms were visually inspected as a last pass, and confidence scores reduced by one if the thermal distribution appeared to be unduly influenced by widely separated records.

We then recalculated CTI for using confidence scores for each species, weighted by their abundance (also $\log(x+1)$ transformed), creating a CTI confidence score for each transect and each site. A mean site confidence score of >2.5 was used as a cut-off for many analyses and figures, as indicated in figure captions. Although a score of 2.5 can be achieved in many ways, this effectively represents at least 75% of the individuals present belonging to species with the maximum confidence score of three.

Thermal guilds. Given few truly subtropical species were identified in this study, and this outcome could potentially result from bias in the distribution of sampling effort towards areas outside of subtropical locations (see Supplementary Information for more detail), we replicated Fig. 2 along a comprehensively sampled latitudinal gradient in Australia. The majority of Australian species are well-sampled across their geographic distributions and numerous sites have been surveyed in subtropical locations in Australia. We divided the RLS data from 968 sites into 10° latitudinal bands along the east coast of Australia (and Papua New Guinea and Solomon Islands) from the equator to 43.7° S, and plotted histograms of thermal distribution midpoints of 1,105 species with a confidence of two or three (Extended Data Fig. 6). These clearly show very few species with midpoints of 23–24 °C, even in the band from 20° S to 30° S where the mean annual SST of sites was 23.97 °C. They also show the intrusion of numerous tropical species in temperate latitudes, particularly for fishes.

Vulnerability predictions. Vulnerability predictions required characterization of the warmest temperatures experienced by species across their range. We reconstructed the thermal distributions for each species using the maximum of the weekly mean SST from all occurrence sites over the 12 weeks before the sampling date, obtaining the 95th percentile of these. We then calculated the difference between this value and the mean of summer temperatures (the mean of the warmest 8 weeks was taken for each year between 2008 and 2014, with the mean of these used). This is analogous to a form of thermal safety margin, although in this case it does not mean a species cannot survive if the summer SST exceeds the 95th percentile, but rather that it has been recorded at very few sites in the combined RLS and GBIF databases at times in which the temperatures exceeded this value.

We re-calculated this value for 10 years and 100 years from present, using rates of SST warming projected by coupled climate models' CMIP5 PCP8.5 scenario, calculated and freely provided by the NOAA Ocean Climate Change Web Portal (http://www.esrl.noaa.gov/psd/ipcc/ocn/). Sea surface temperature anomaly (difference in the mean climate in the future time period, 2050–2099, compared to the historical reference period, 1956–2005) was selected as the statistic representing the average of 25 models, interpolated to a 1° latitude by 1° longitude grid and matched to each RLS site. Summer SST was predicted for each RLS site for 10 and 100 year time periods using these values. Vulnerability was then estimated as the proportion of all species (fishes and invertebrates) recorded on each RLS survey that is expected to exceed the 95th percentile, based on the predicted SST at that site. This component of analyses did not incorporate abundance data, as the goal was to assess local species loss, rather than loss of individuals. Weighting by abundance had little influence on conclusions, however.

Confidence scores were also recalculated without abundance (and thus represent the mean confidence of species present), and sites with confidence scores <2.5 were excluded from calculation of ecoregion means for all ecoregions with three or more sites with confidence >2.5. Twenty-one of 81 ecoregions had fewer than three sites with confidence >2.5 with which to calculate means, so low confidence sites were included in means for these ecoregions. The effect of this is conservative, theoretically reducing thermal bias (see Supplementary Information), but the rationale was that ecoregion means would be more accurate through their inclusion than if heavily weighted by few sites. To provide an additional cut-off for ecoregions in which the overall mean confidence was still low, we excluded ecoregions with

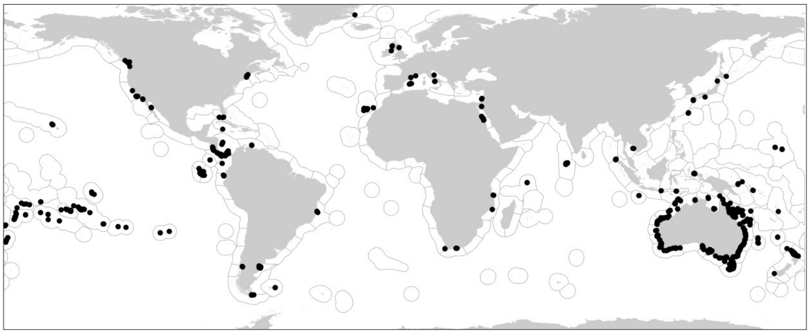
mean confidence <1.75. This resulted in the exclusion of six ecoregions (North and East Barents Sea, Oyashio Current, Agulhas Bank, Sea of Japan/East Sea, Gulf of Maine/Bay of Fundy, Malvinas/Falklands).

To explore the contributions of warming rates and thermal bias to vulnerability predictions, we also recalculated CTI as the mean 95th percentiles of fish and invertebrate species recorded on transects (CTI_{max}) and thermal bias ($TBias_{max}$) as the difference between site-level CTI_{max} and mean summer SST. $TBias_{max}$ can therefore be considered the sensitivity component of the vulnerability predictions, based on recent mean summer SST and not accounting for warming rates (exposure). We applied GAMMs to assess vulnerability scores as a function of $TBias_{max}$ and warming rates, with ecoregion as a random factor (Extended Data Table 2).

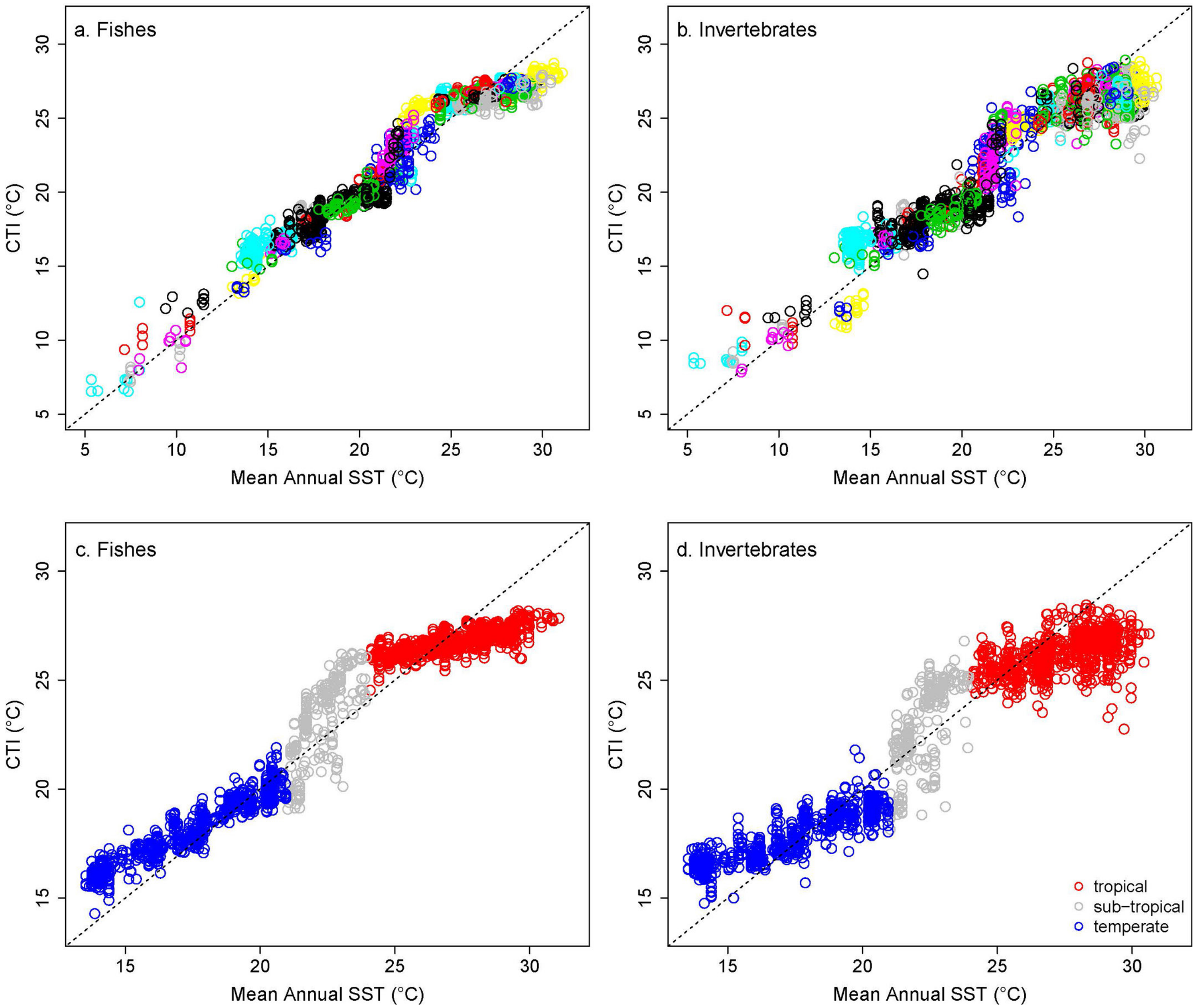
Conclusions are robust to the warming data used, with qualitatively similar results using historical warming data from another source[8], instead of future predictions (site warming rates in °C per decade taken from http://www.coastalwarming.com/data.html), and ecoregion mean vulnerability scores changing very little when the 99th percentile of species' thermal distributions were used instead of the 95th percentile, even for 2115 predictions (Pearson correlation $= 0.97$, $P < 0.01$).

Data reporting. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

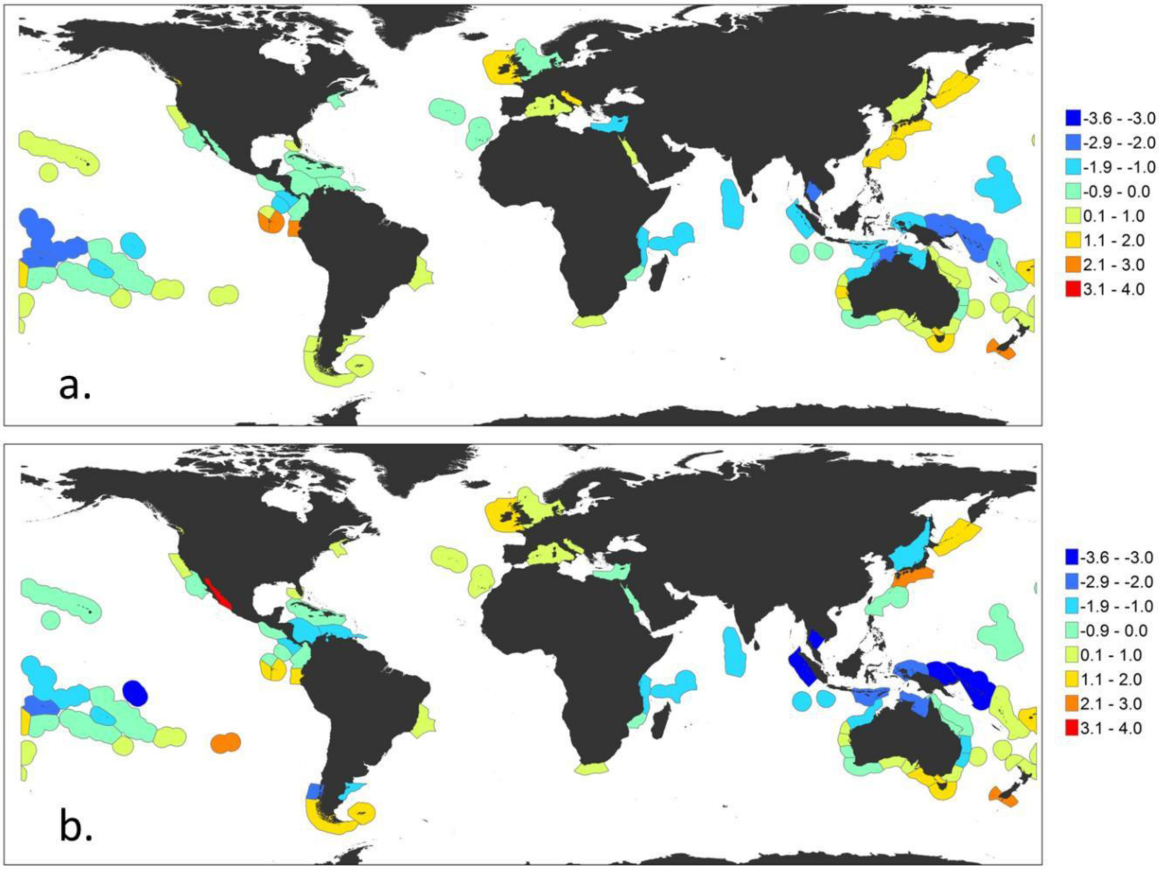
44. Bates, A. E. *et al.* Distinguishing geographical range shifts from artefacts of detectability and sampling effort. *Divers. Distrib.* **21,** 13–22 (2015).



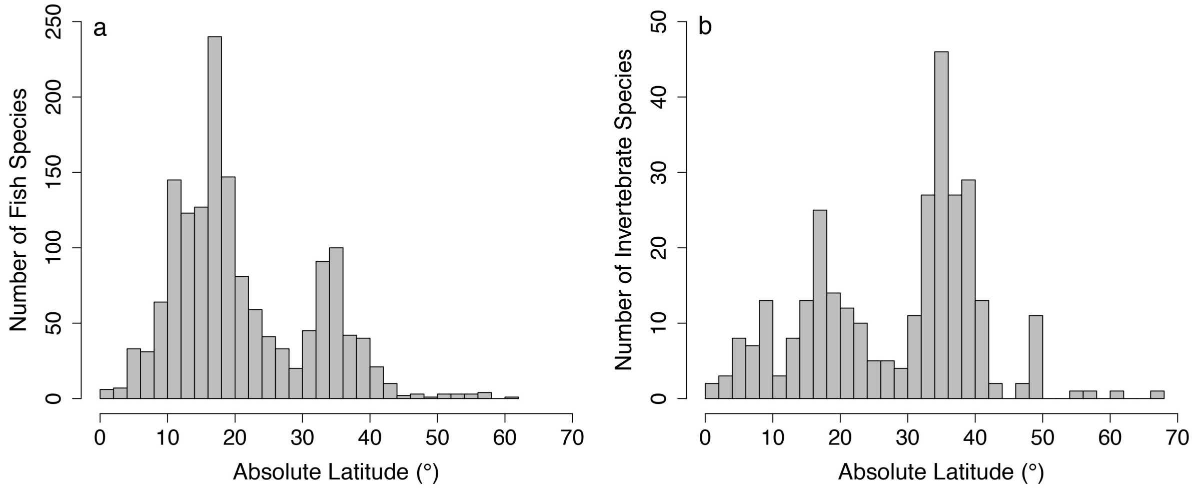
Extended Data Figure 1 | Sites used in analyses at which fish and invertebrate communities were surveyed by the Reef Life Survey program. Numerous points are overlapping and hidden ($n = 2{,}447$). Ecoregion boundaries are shown in grey lines.



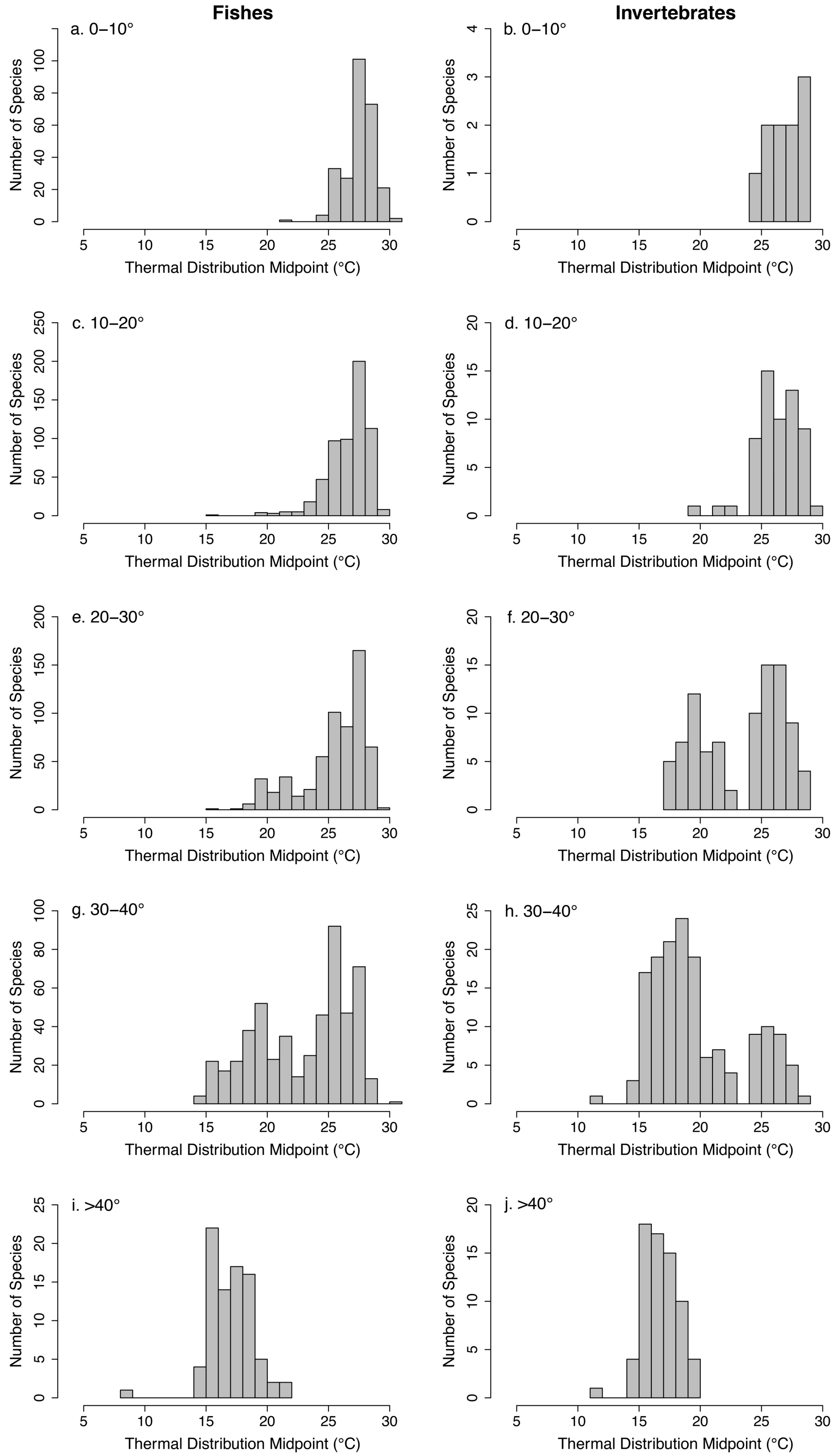
Extended Data Figure 2 | Community temperature index values for reef fishes and invertebrates against mean annual sea surface temperature. a–d, CTI calculated using abundance-weighted fish (**a**) and invertebrate (**b**) data, and including sites at which mean CTI confidence scores were less than 2.5 ($n = 2,447$ and 2,383 for fishes and invertebrates, respectively). Sites are colour-coded by ecoregion to help distinguish spatial patterns, but as a result of numerous ecoregions ($n = 81$), many ecoregion colours are similar. CTI calculated using presence-only fish (**c**) and invertebrate (**d**) data, and excluding sites with confidence scores <2.5 ($n = 2,188$ and 1,812 for fishes and invertebrates, respectively). Dotted lines have a slope of one, plotted for comparison with data.



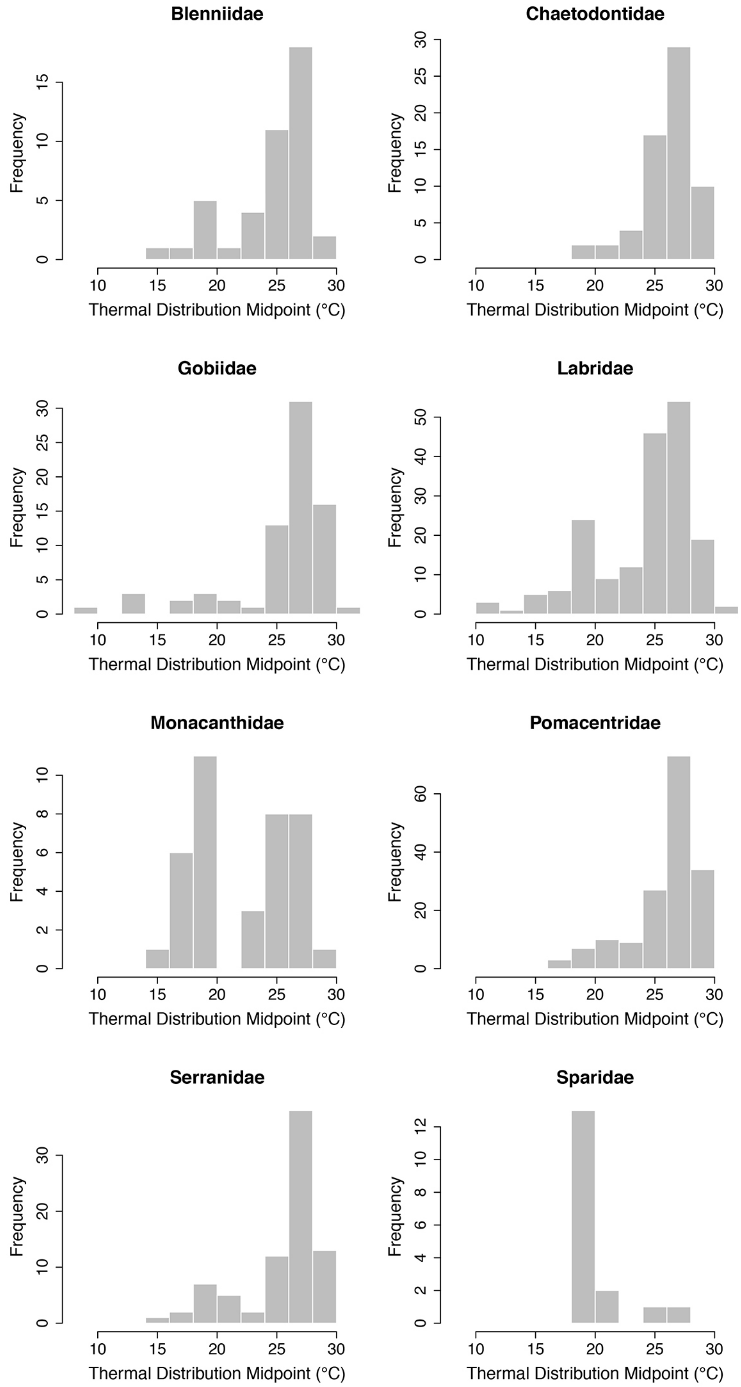
Extended Data Figure 3 | Global distribution of reef fish and invertebrate community thermal bias. a, **b**, Community thermal bias (°C) is the difference in abundance-weighted CTI from local long-term mean annual sea surface temperature. Positive regions (warm colours) encompass ecological communities with a predominance of individuals with warmer thermal affinity than mean local sea temperatures. Colours are scaled to the mean thermal bias of sites surveyed within each ecoregion (see Extended Data Table 1 for sample sizes). Only ecoregions with sites that were surveyed are included.



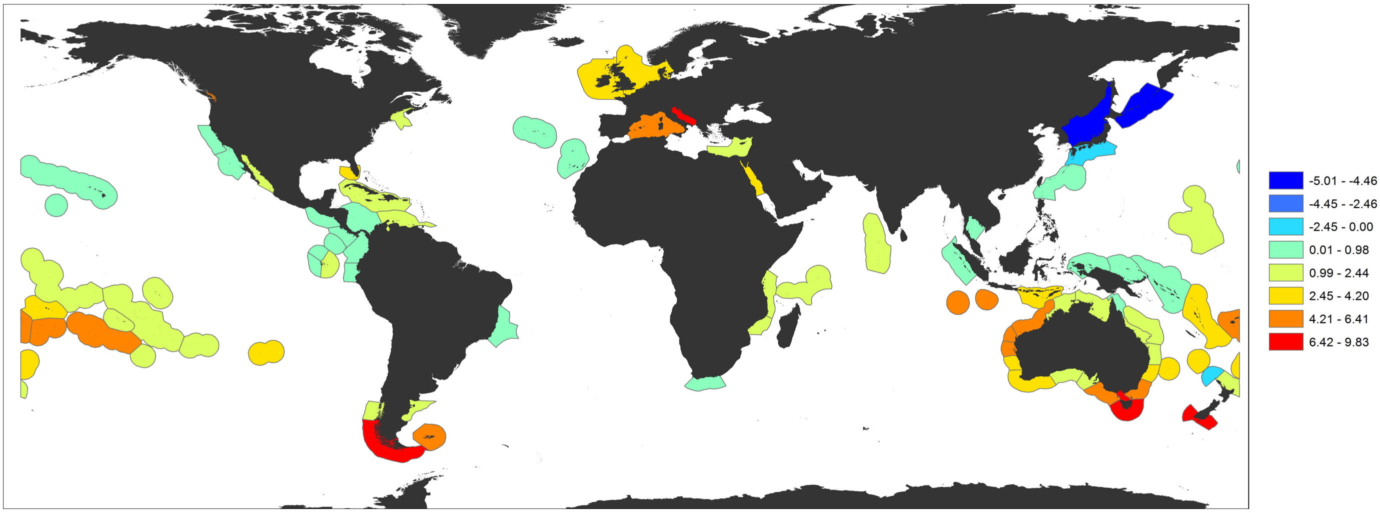
Extended Data Figure 4 | Frequency distribution of fish and invertebrate species' latitudinal range midpoints. a, **b**, Species for which confidence in thermal distribution midpoints (and therefore geographical distribution midpoints) was low are excluded (see Methods).



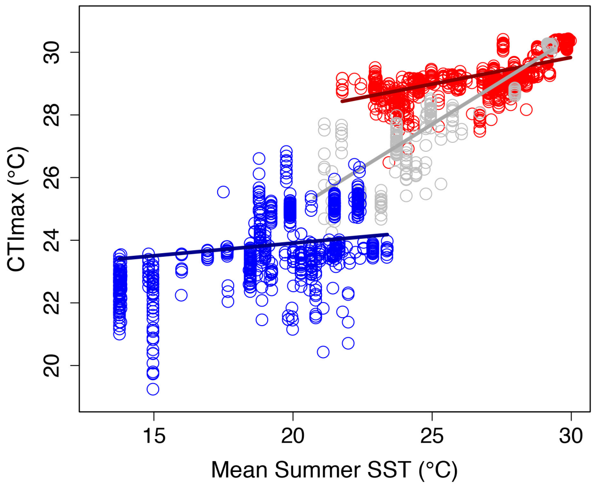
Extended Data Figure 5 | Frequency distribution of fish (left) and invertebrate (right) species' thermal distribution midpoints in 10° latitudinal bands from Papua New Guinea and down eastern Australia (rows). a–j, Note *y* axes are on different scales and only species with confidence scores of two and three are included (see Methods).



Extended Data Figure 6 | Frequency distribution of thermal distribution midpoints of species in major fish families spanning temperate and tropical zones. Note *y* axes are on different scales and only species with confidence scores of two and three are included.



Extended Data Figure 7 | Global distribution of $TBias_{max}$ of reef faunal communities. $TBias_{max}$ is calculated as the difference between CTI_{max} (using the 95th percentiles of species' thermal distributions and presence data) and mean summer SST. Colours are scaled to the mean $TBias_{max}$ of sites surveyed within each ecoregion (see Extended Data Table 1 for sample sizes). Only ecoregions in which quantitative surveys were undertaken are included.



Extended Data Figure 8 | The CTI_{max} (mean 95th percentile of species thermal distributions) for reef faunal communities across temperate (blue), tropical (red) and subtropical (grey) sites. SST data are means of the warmest 8 weeks of the year over the survey period (2008–2014). Points represent the surveyed community of fishes and invertebrates at each site ($n = 2{,}091$, only confidence scores >2.5). Regression lines are fitted to the maximum values within each ecoregion, with separate regressions fitted for sites categorised from Fig. 1 as temperate, tropical and subtropical.

Extended Data Table 1 | Ecoregion means, sample sizes and vulnerability predictions

ECOREGION	Group (Fig 1)	# sites (ED Fig 3a)	# sites (ED Fig 3b)	# sites (Fig 3, ED Fig 7)	$TBias_{max}$ (ED Fig 7)	Vulnerability 2025 (Fig 3b)	Vulnerability 2115 (Fig 3d)
Adriatic Sea	TE	1 (1)*	(2)*	(2)*	7.08	0.00	0.32
Agulhas Bank	TE	(11)*	(11)*				
Arnhem Coast to Gulf of Carpenteria	TE	12 (0)	9 (3)*	12	1.20	0.24	0.85
Azores Canaries Madeira	TE	6 (57)	32 (31)	12 (51)	0.76	0.15	0.99
Bassian	TE-SP	236 (5)	234 (7)	237 (4)	7.06	0.00	0.03
Bight of Sofala/Swamp Coast	TR	3 (0)	2 (1)	3	1.28	0.06	0.99
Bismarck Sea	TR	9 (1)	1 (0)	9 (1)	0.49	0.17	1.00
Bonaparte Coast	TR	15 (1)	8 (8)*	14 (2)	1.67	0.18	0.58
Cape Howe	TE	183 (0)	181 (1)	180 (1)	5.27	0.00	0.03
Celtic Seas	SP	3 (6)*	2 (7)*	1 (8)*	3.24	0.52	0.58
Central and Southern Great Barrier Reef	TR	62 (0)	38 (17)	62	1.28	0.04	0.83
Central Kuroshio Current	ST	(9)*	4 (5)*	(9)*	-0.97	0.44	0.94
Channels and Fjords of Southern Chile	SP	(8)*	(8)*	(8)*	6.54	0.21	0.21
Chiapas-Nicaragua	TR	14 (0)	9 (5)*	14	0.57	0.05	1.00
Chiloense	SP	(9)*	(9)*	(9)*	1.70	0.00	0.58
Cocos Islands	TR	23 (1)	22 (1)	23	0.79	0.00	1.00
Cocos-Keeling/Christmas Island	TR	15 (0)	11 (2)	15	6.15	0.03	0.03
Coral Sea	TR	154 (0)	136 (14)	154	1.34	0.04	0.79
Cortezian	TR	(8)*	8 (0)	(8)*	2.06	0.24	0.42
East African Coral Coast	TR	9 (0)	9 (0)	9	1.66	0.00	0.96
Easter Island	ST	2 (15)*	17 (0)	3 (14)	3.32	0.00	0.26
Eastern Brazil	TR	(10)*	1 (9)*	(10)*	0.15	0.67	1.00
Eastern Galapagos Islands	ST	36 (0)	36 (0)	31	1.00	0.01	1.00
Exmouth to Broome	TR	176 (3)	141 (31)	176 (3)	4.39	0.02	0.05
Fiji Islands	TR	9 (0)	9 (0)	9	5.86	0.01	0.02
Floridian	TR	17 (0)	1 (14)*	16 (1)	2.67	0.00	0.06
Great Australian Bight	TE	3 (0)	3 (0)	3	1.27	0.07	1.00
Greater Antilles	TR	1 (0)	(1)*	1	1.16	0.01	1.00
Guayaquil	ST	16 (0)	16 (0)	15	0.96	0.01	1.00
Gulf of Maine/Bay of Fundy	SP	1 (7)*	(8)*				
Gulf of Thailand	TR	7 (0)	7 (0)	5 (2)	0.11	0.63	1.00
Hawaii	TR	2 (7)*	9 (0)	1 (8)*	0.73	0.21	0.79
Houtman	TE	32 (0)	29 (3)	32	3.40	0.00	0.36
Kermadec Island	TE	14 (0)	8 (6)*	13 (1)	2.88	0.22	0.24
Leeuwin	TE	69 (7)	63 (12)	71 (5)	2.86	0.03	0.25
Lesser Sunda	TR	11 (0)	10 (1)	11	3.69	0.01	0.04
Levantine Sea	ST	6 (0)	(5)*	2 (4)*	2.69	0.22	0.46
Lord Howe and Norfolk Islands	ST	94 (3)	83 (14)	91 (6)	3.20	0.01	0.17
Maldives	TR	12 (0)	10 (2)	12	1.08	0.02	1.00
Malvinas/Falklands	SP	(5)*	(5)*				
Manning-Hawkesbury	ST	141 (0)	133 (8)	140 (1)	3.16	0.00	0.10
Marquesas	TR	(7)*	6 (1)	3 (4)	1.18	0.08	1.00
Marshall Islands	TR	16 (0)	11 (2)	11	1.69	0.04	0.72
Nicoya	TR	93 (0)	89 (4)	89 (1)	0.57	0.01	1.00
Ningaloo	TR	30 (0)	30 (0)	30	5.66	0.04	0.05
North and East Barents Sea	SP	(2)*	(2)*				
North Patagonian Gulfs	TE-SP	(13)*	(13)*	(13)*	2.14	0.00	0.44
North Sea	SP	6 (0)	1 (5)*	(6)*	3.43	0.49	0.67
Northeastern New Zealand	TE	75 (28)	30 (73)	31 (71)	1.86	0.32	0.52
Northern and Central Red Sea	TR	13 (4)	6 (7)*	11 (6)	4.24	0.14	0.22
Northern California	TE-SP	(8)*	(8)*	(8)*	0.49	0.67	1.00
Northern Galapagos Islands	TR	11 (0)	10 (1)	9	0.93	0.00	1.00
Oyashio Current	SP	(4)*	(4)*				
Panama Bight	TR	40 (0)	39 (1)	40	0.47	0.07	1.00
Papua	TR	30 (3)	21 (11)	28 (3)	0.64	0.02	1.00
Phoenix/Tokelau/Northern Cook Islands	TR	12 (0)	6 (6)*	12	2.38	0.01	0.62
Puget Trough/Georgia Basin	SP	(8)*	(8)*	(8)*	4.24	0.15	0.38
Rapa-Pitcairn	ST	5 (0)	5 (0)	5	1.86	0.01	0.77
Samoa Islands	TR	25 (0)	19 (6)	25	2.98	0.03	0.26
Sea of Japan/East Sea	TE-SP	(6)*	(6)*				
Seychelles	TR	12 (0)	12 (0)	12	2.22	0.00	0.57
Shark Bay	ST	6 (0)	4 (1)	6	5.56	0.08	0.13
Society Islands	TR	17 (0)	8 (7)*	17	2.32	0.00	0.55
Solomon Archipelago	TR	5 (0)	5 (0)	5	0.96	0.04	1.00
South Australian Gulfs	TE	71 (0)	70 (1)	71	1.63	0.07	0.72
South Kuroshio	TR	8 (0)	8 (0)	8	0.84	0.18	1.00
South New Zealand	SP	(1)*	(1)*	(1)*	9.83	0.11	0.11
Southern California Bight	TE	(14)*	(14)*	(14)*	0.81	0.23	0.97
Southern Caribbean	TR	14 (0)	10 (4)*	14	1.35	0.00	1.00
Southern Cook/Austral Islands	TR	15 (0)	13 (2)	15	5.10	0.01	0.01
Southwestern Caribbean	TR	22 (0)	9 (12)*	22	0.33	0.55	1.00
Three Kings-North Cape	TE	6 (0)	(6)*	4 (2)	-0.31	0.66	0.72
Tonga Islands	TR	31 (0)	22 (8)	31	5.46	0.05	0.07
Torres Strait Northern Great Barrier Reef	TR	26 (0)	21 (3)	26	0.90	0.07	1.00
Tuamotus	TR	53 (0)	36 (7)	53	1.20	0.03	0.97
Tweed-Moreton	ST	39 (0)	33 (5)	39	2.12	0.16	0.54
Vanuatu	TR	1 (0)	1 (0)	1	3.75	0.01	0.01
Western Bassian	TE-SP	10 (0)	10 (0)	10	4.60	0.03	0.07
Western Galapagos Islands	ST	30 (1)	31 (0)	30 (1)	0.88	0.00	1.00
Western Mediterranean	TE	28 (0)	26 (2)	26 (2)	6.29	0.04	0.14
Western Sumatra	TR	30 (0)	27 (0)	30	0.61	0.13	1.00

The number of sites used in figures is the number of sites with confidence >2.5, with number of sites with confidence <2.5 shown in brackets. An asterisk indicates that sites with confidence <2.5 were included in calculations of ecoregion means. Group identifies whether fauna surveyed at sites within the ecoregion can be classified as temperate (TE), tropical (TR), subtropical (ST), subpolar (SP), and temperate-subpolar transition (TE-SP) on the basis of CTI.

Extended Data Table 2 | GAMM results

	coefficient	standard error	t-value	P-value
2025				
Intercept	0.080	0.137	0.586	0.558
Warming rate	-0.138	0.592	-0.233	0.816
$TBias_{max}$				<0.001*
2115				
Intercept	0.204	0.137	1.490	0.136
Warming rate	1.180	0.591	1.999	0.046*
$TBias_{max}$				<0.001*

Results for Fig. 3b and d. Proportion of species loss predicted by 2025 and 2115 as a function of warming rate and $TBias_{max}$. $n = 2{,}091$.

ARTICLE

doi:10.1038/nature16071

Brain tumour cells interconnect to a functional and resistant network

Matthias Osswald[1,2], Erik Jung[1,2*], Felix Sahm[3,4*], Gergely Solecki[1,2*], Varun Venkataramani[5], Jonas Blaes[1,2], Sophie Weil[1,2], Heinz Horstmann[5], Benedikt Wiestler[1,2,6], Mustafa Syed[1,2], Lulu Huang[1,2], Miriam Ratliff[2,7], Kianush Karimian Jazi[1,2], Felix T. Kurz[8], Torsten Schmenger[1,2], Dieter Lemke[1,2], Miriam Gömmel[1,2], Martin Pauli[9], Yunxiang Liao[1,2], Peter Häring[10], Stefan Pusch[3,4], Verena Herl[11], Christian Steinhäuser[11], Damir Krunic[12], Mostafa Jarahian[13], Hrvoje Miletic[14], Anna S. Berghoff[15], Oliver Griesbeck[16], Georgios Kalamakis[17], Olga Garaschuk[17], Matthias Preusser[18], Samuel Weiss[19,20,21], Haikun Liu[22], Sabine Heiland[8], Michael Platten[1,23], Peter E. Huber[24,25], Thomas Kuner[5], Andreas von Deimling[3,4], Wolfgang Wick[1,2] & Frank Winkler[1,2]

Astrocytic brain tumours, including glioblastomas, are incurable neoplasms characterized by diffusely infiltrative growth. Here we show that many tumour cells in astrocytomas extend ultra-long membrane protrusions, and use these distinct tumour microtubes as routes for brain invasion, proliferation, and to interconnect over long distances. The resulting network allows multicellular communication through microtube-associated gap junctions. When damage to the network occurred, tumour microtubes were used for repair. Moreover, the microtube-connected astrocytoma cells, but not those remaining unconnected throughout tumour progression, were protected from cell death inflicted by radiotherapy. The neuronal growth-associated protein 43 was important for microtube formation and function, and drove microtube-dependent tumour cell invasion, proliferation, interconnection, and radioresistance. Oligodendroglial brain tumours were deficient in this mechanism. In summary, astrocytomas can develop functional multicellular network structures. Disconnection of astrocytoma cells by targeting their tumour microtubes emerges as a new principle to reduce the treatment resistance of this disease.

Astrocytomas (World Health Organisation grades II, III and IV; grade IV are called glioblastomas) are prototypical examples for highly invasive tumours that diffusely colonize their host organ[1], which ultimately leads to neurological dysfunction and death despite intensive radio- and chemotherapy. Oligodendroglioma is another glioma type that shares many molecular features like frequent isocitrate dehydrogenase (*IDH1* and *IDH2*) mutations[2,3], but is less invasive and far more vulnerable to therapeutic intervention than astrocytomas. A codeletion of the chromosomal parts 1p and 19q is characteristic for oligodendrogliomas, but absent in astrocytomas[2–4]. This codeletion allows for molecular subgrouping[2–4], and is associated with a high responsiveness of oligodendrogliomas to radiochemotherapy, leading to marked long-term survival benefits[5,6]. The reason for that remained unclear, just as the specific mechanism(s) of resistance in 1p/19q intact astrocytomas.

Here we describe the discovery of a direct anatomical connection between astrocytoma cells, with relevance for tumour functionality and resistance. Similar cell–cell connections by membrane tubes have been first described in *Drosophila* development[7]. They can play a role in the transport of organelles and proteins[8,9], spread of infectious particles[10], stem cell signalling[11], and functional cell–cell coupling[12–14]. Studied mostly *in vitro* so far, these tubes have received many names, including membrane nanotubes, tunnelling nanotubes, or cytonemes. However, the exact function(s) of membrane tube connections in mammalian tissues and in tumour biology remained unresolved[15].

Membrane tubes in glioma progression

To study the occurrence and dynamics of membrane tube protrusions in mammalian tumours, we followed gliomas growing in the mouse brain by *in vivo* multiphoton laser-scanning microscopy (MPLSM) down to a depth of 750 µm (ref. 16), for up to one year. After transplantation of patient-derived glioblastoma cell lines ($n = 6$) that were kept under serum-free, stem-like conditions[17] (GBMSCs; non-codeleted for 1p/19q, and *IDH* wild-type; Extended Data Fig. 1a–k), many tumour cells formed ultra-long cellular protrusions.

[1]Neurology Clinic and National Center for Tumor Diseases, University Hospital Heidelberg, INF 400, 69120 Heidelberg, Germany. [2]Clinical Cooperation Unit Neurooncology, German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [3]Department of Neuropathology, Institute of Pathology, Ruprecht-Karls University Heidelberg, INF 224, 69120 Heidelberg, Germany. [4]Clinical Cooperation Unit Neuropathology, German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), INF 224, 69120 Heidelberg, Germany. [5]Department of Functional Neuroanatomy, Institute of Anatomy and Cell Biology, Heidelberg University, INF 307, 69120 Heidelberg, Germany. [6]Department of Diagnostic and Interventional Neuroradiology, Klinikum rechts der Isar der Technischen Universität München, 81675 Munich, Germany. [7]Neurosurgery Clinic, University Hospital Heidelberg, INF 400, 69120 Heidelberg, Germany. [8]Department of Neuroradiology, University Hospital Heidelberg, INF 400, 69120 Heidelberg, Germany. [9]Department of Neurophysiology, Institute of Physiology, University of Würzburg, 97070 Würzburg, Germany. [10]Department of Medical Physics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [11]Institute of Cellular Neurosciences, Medical Faculty, University of Bonn, Sigmund-Freud-Strasse 25, 53105 Bonn, Germany. [12]Light Microscopy Facility, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [13]Department of Translational Immunology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [14]Department of Biomedicine, University of Bergen, Jonas Lies vei 91, 5009 Bergen, Norway. [15]Institute of Neurology, Medical University of Vienna, Vienna, Austria; Comprehensive Cancer Center, CNS Unit, Medical University of Vienna, 1090 Vienna, Austria. [16]Tools For Bio-Imaging, Max-Planck-Institute of Neurobiology, 82152 Martinsried, Germany. [17]Institute of Physiology II, Eberhard Karls University of Tübingen, 72074 Tübingen, Germany. [18]Department of Medicine I, Medical University of Vienna, Vienna, Austria; Comprehensive Cancer Center, CNS Unit, Medical University of Vienna, 1090 Vienna, Austria. [19]Hotchkiss Brain Institute, Faculty of Medicine, University of Calgary, Calgary, Alberta T2N 4N1, Canada. [20]Department of Cell Biology and Anatomy, Faculty of Medicine, University of Calgary, Calgary, Alberta T2N 4Z6, Canada. [21]Clark Smith Brain Tumor Research Centre, Southern Alberta Cancer Research Institute, Faculty of Medicine, University of Calgary, Calgary, Alberta, T2N 4N1, Canada. [22]Helmholtz Young Investigator Group, Normal and Neoplastic CNS Stem Cells, DKFZ-ZMBH Alliance, German Cancer Research Center (DKFZ), INF 280, 69120 Heidelberg, Germany. [23]Clinical Cooperation Unit Neuroimmunology and Brain Tumor Immunology, German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [24]CCU Molecular and Radiation Oncology, German Cancer Research Center (DKFZ), INF 280, 69120 Heidelberg, Germany. [25]Department of Radiation Oncology, University Hospital Heidelberg, 69120 Heidelberg, Germany.
*These authors contributed equally to this work.

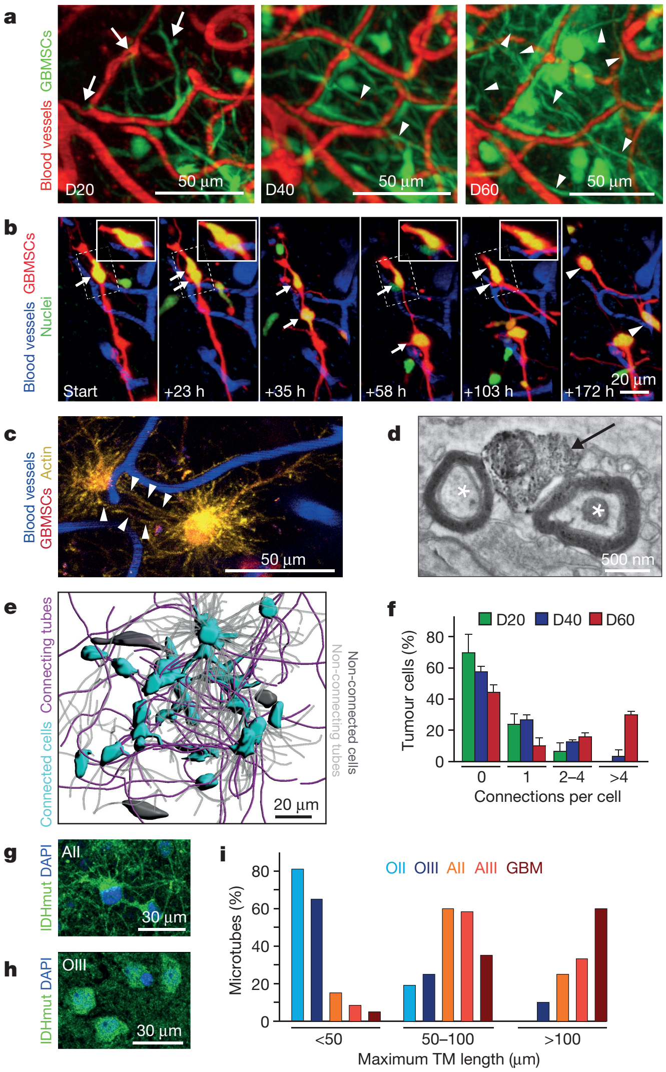


Figure 1 | Distinct membrane microtubes of brain tumour cells. **a**, *In vivo* MPLSM maximum intensity projections (MIPs) of S24 GBMSCs growing in the mouse brain over 60 days (D). Arrows, thin cellular protrusions extending into the normal brain; arrowheads, long intratumoral protrusions. **b**, Travel of nuclei (arrows, arrowheads) after nuclear division (at 23 and 103 h) in cellular protrusions of S24 GBMSCs. 3D images. **c**, Protrusions are actin-rich, and interconnect single tumour cells (arrowheads, S24 Lifeact-YFP, MIP). **d**, Scanning electron microscopy (SEM) image of a S24 GBMSC membrane microtube (arrow, identified by GFP photo-oxidation) in the mouse brain. Asterisks, axons. **e**, 3D rendering (*z* dimension 90 μm) of membrane microtubes interconnecting single S24 GBMSCs. Intercellular (connecting) and non-connecting tubes, and connected and unconnected tumour cells are shown colour-coded. **f**, Number of membrane microtubes per S24 tumour cell that connect this cell to another tumour cell ($n = 141$–437 cells in $n = 3$ mice). **g**, **h**, Representative confocal *IDH1R132H* mutation-specific immunofluorescence images of a human astrocytoma grade II (AII, **g**) and oligodendroglioma grade III (OIII, **h**). **i**, Maximum length of $IDH1^{R132H}$-positive microtubular structures in human oligodendrogliomas (O, 1p/19q codeleted) grade II, III; and astrocytomas (A, 1p/19q non-codeleted) grade II, III and IV (glioblastoma, GBM); $n = 20$–24 patients per tumour entity, $n = 105$ total. **a**–**c** and **e**–**f**, *in vivo* MPLSM. Error bars show s.d.

These protrusions infiltrated the normal brain at the invasive front (Fig. 1a), where astrocytoma cells extended and retracted them in a scanning mode (Supplementary Video 1). Protrusion tips were highly dynamic (Extended Data Fig. 1l), similar to neuronal growth cones during development[18]. When tumours progressed, the number of cellular protrusions increased further, some exceeding 500 μm in length (Extended Data Fig. 2a). The resulting membrane tubes where used as tracks for travel of cell nuclei, for example, after mitosis (Fig. 1b; speed of travelling nuclei: 66.42 ± 36.25 μm per day, $n = 16$ nuclei in $n = 6$ mice.). These data suggest that membrane tube formation is a novel means of tumour dissemination, adding to the known strategies[19].

All membrane tubes were actin-rich (Fig. 1c), which is also typical for most membrane nanotubes[8]. Moreover, live imaging and immunohistochemistry revealed that they were indeed enclosed by a continuous cell membrane, and positive for myosin IIa, microtubules, and protein disulphide isomerase; partly positive for β-catenin, β-parvin, and the astrocytic marker GFAP; but largely negative for N-cadherin, myosin X, and the neuronal marker β-tubulin III (Extended Data Fig. 2b, c). Together, these data indicate that these membrane tubes have a unique composition and a potent motility machinery. Dendritic arborization was frequent, with more dynamic thin membrane tubes originating from more stable, thicker ones (Fig. 1c, Extended Data Fig. 3a, b). To allow ultrastructural analysis of these thin tumour cell-derived tubes in the mouse brain by electron microscopy, photo-oxidation of brain sections was performed. This resulted in dark precipitates within the green fluorescent protein (GFP)-expressing astrocytoma cell tubes (Extended Data Fig. 3c). Serial-section scanning electron microscopy (3D SEM) revealed that the cell membrane-enclosed tubes had a mean cross-sectional area of 1.57 ± 0.33 μm^2 ($n = 6$, Extended Data Fig. 3d), and contained mitochondria and microvesicles (Fig. 1d), suggesting that there is local ATP production and vesicle trafficking in the tubes. Interestingly, mitochondria travelled quickly in these tubes (Extended Data Fig. 3e). Furthermore, a relevant number of membrane microtubes were following axons in the brain (19.6% of $n = 51$; Fig. 1d, Extended Data Fig. 3f), which are known leading structures for tumour cell dissemination in astrocytomas[20].

In vivo imaging of membrane tube development over time revealed that an increasing number started at one and ended at another astrocytoma cell, creating a multicellular anatomical network (Fig. 1e, f; Supplementary Video 2a, b). Abundant intercellular membrane tubes were also found in a genetic astrocytoma model[21] (Extended Data Fig. 3g, Supplementary Video 2c). Intercellular membrane tubes were in part a result of cell division, with enduring stable contact of daughter cells over long distances (Fig. 1b), but also of mating of non-related astrocytoma cells (Extended Data Fig. 4a–g). A small proportion of membrane tube-bearing astrocytoma cells maintained quiescent for months, often in a perivascular niche that has been associated with glioma cell stemness[22] (Extended Data Fig. 4h, i).

The intercellular position of many astrocytoma membrane tubes, together with their high content of F-actin, is reminiscent of membrane nanotubes[15]; however, the membrane nanotubes reported so far had a width of below 1 μm; a length of usually tens, rarely a few hundreds of μm; and documented life time of less than 60 min. These differences led us to propose the new term “tumour microtubes”, or TMs, for the discovered ultra-long, long-lived, and thicker membrane extensions of astrocytoma cells.

TMs characterize human astrocytomas

To investigate whether TMs are also characteristic for human brain tumours, we stained resected WHO grade II–IV gliomas with $IDH1^{R132H}$ mutations using a mutation-specific antibody[23]. This allowed us to unambiguously detect tumour-cell-derived membrane tubes in the filament-rich brain parenchyma. Like in the astrocytoma mouse models, TMs were abundant in patient tissue (Fig. 1g): 63% of astrocytoma cells had intercellular TMs ($n = 196$ cells in 100 μm thick sections of $n = 8$ WHO grade II–III tumours without 1p/19q codeletion; Supplementary Video 3). In contrast, only 0.7% of oligodendroglioma cells in human tumour samples had intercellular TMs (Fig. 1h; $n = 150$ cells from $n = 3$ oligodendrogliomas with 1p/19q codeletion), and TMs were also rare in patient-derived oligodendroglioma cells that formed tumours in mice (Extended Data Fig. 5a, b). Further analysis of 105 human gliomas revealed that TM formation was highly influenced by tumour type and grade, with a marked positive correlation of TM

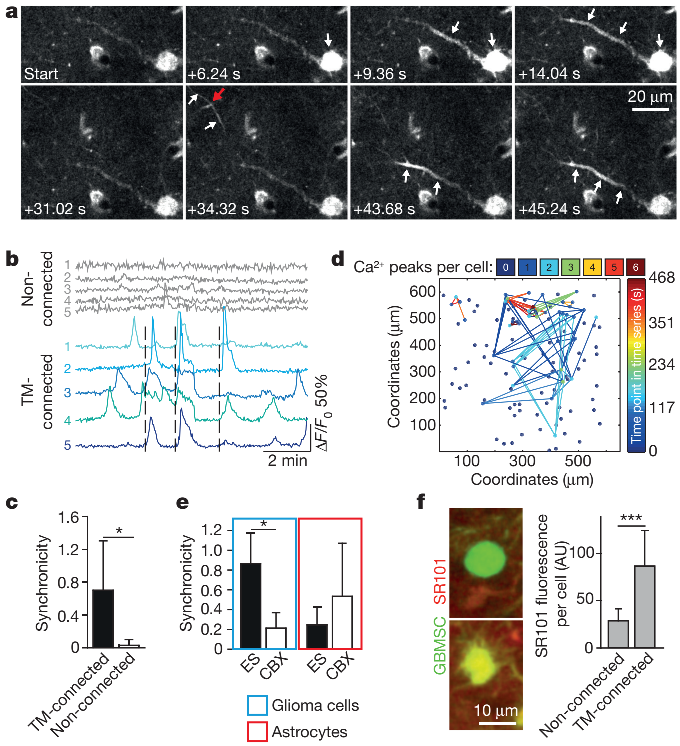


Figure 2 | TM-connections allow communication in multicellular networks. **a**, Time-lapse image of calcium waves (Rhod2-AM) travelling along TMs of GBMSCs (bidirectional; arrows). Red arrow, crossing of two TMs, with simultaneous calcium peak. **b**, Calcium transients ($\Delta F/F_0$, Rhod-2AM) of non-connected GMBSCs (grey) versus TM-connected cells (blue); broken lines mark synchronous calcium transients. **c**, Synchronicity (see Methods section) of calcium peaks in GBMSCs, shown for TM-connected versus non-TM-connected tumour cells ($n = 40$ versus 43 cells in $n = 3$ mice per condition; *t*-tests). **d**, Representative heat map of calcium transients between GBMSCs. **e**, Synchronicity of calcium peaks during brain superfusion with extracellular saline (ES, control) versus 100 µM carbenoxolone (CBX) in GBMSCs (blue box) and normal brain astrocytes (red box); $n = 3$ mice per group; *t*-tests. **f**, SR101 uptake in a non-TM-connected GBMSC (upper image) and a TM-connected one (lower image). Right, corresponding quantification (AU, arbitrary units); $n = 55$ cells in $n = 3$ mice per condition; Mann–Whitney test. All images and analyses, *in vivo* MPLSM. GBMSCs, S24 line. Error bars show s.d. $*P < 0.05$, $***P < 0.001$.

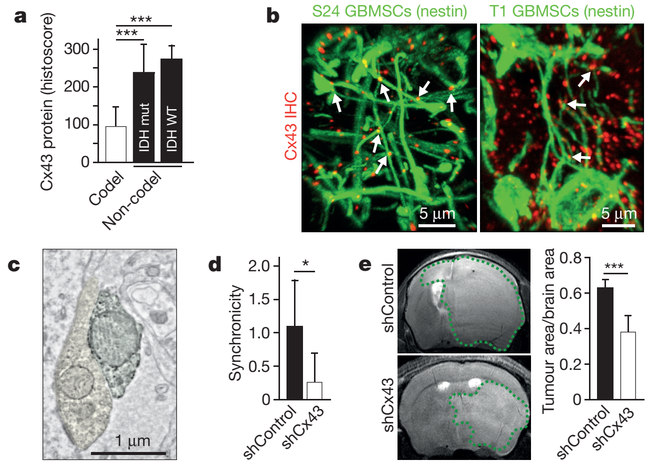


Figure 3 | Connexin 43 gap junctions connect TMs. **a**, Quantification of Cx43 protein expression detected by immunohistochemistry in 1p/19 non-codeleted (*IDH* wild-type and *IDH* mutated) versus codeleted human gliomas ($n = 8$ each, ANOVA, Tukey's post-hoc test). **b**, Nestin and Cx43 double-immunofluorescence in S24 and T1 GBMSC tumours (arrows: Cx43-positive TM-interconnections). **c**, SEM image of a direct membrane contact of 2 TMs in the mouse brain. **d**, Synchronicity of calcium peaks in S24 shControl versus shCx43 cells ($n = 3$ mice/condition; Mann–Whitney test). **e**, T2 MRI images of shControl versus shCx43 tumours, 72 days after implantation. Quantifications of $n = 6$ animals per group (*t*-test). **d**, *In vivo* MPLSM. Error bars show s.d. $*P < 0.05$; $***P < 0.001$.

length and unfavourable prognosis. For example, bona fide TMs of a minimum of 50 µm length in standard thin sections were detectable in only 19% of WHO grade II oligodendrogliomas, but 93% of WHO grade IV astrocytomas (= glioblastomas) (Fig. 1i). In astrocytomas, TMs were even detected in the contralateral brain hemisphere (Extended Data Fig. 5c), and also in *IDH* wild-type tumours (Extended Data Fig. 5d). The 1p/19q status better predicted TM occurrence than morphological glioma classification (Extended Data Fig. 5e–g).

A communicating network

Intercellular calcium waves (ICWs) can coordinate the activity of individual cells in multicellular networks, which includes astrocytes of the normal brain[24,25], neurons[26], and radial glia cells during central nervous system development[27]. We observed extensive and long-range ICWs, involving many astrocytoma cells in various tumour regions (Supplementary Video 4). ICWs were propagated along TMs in both directions (Fig. 2a, Extended Data Fig. 6a). Further analysis confirmed that ICWs, measured by synchronicity of calcium fluctuations, were largely restricted to astrocytoma cells with detectable TM connections (Fig. 2b, c; Extended Data Fig. 6b), allowing communication of individual cells in a reproducible pattern (Fig. 2b, d; Extended Data Fig. 6c, d).

For membrane nanotubes, intercellular connections have been reported *in vitro* to either be open ended[8], or separated by gap junctions[14], the latter as a prerequisite for ICW propagation[15]. Indeed, pharmacological gap junction blockade reduced the frequency (Extended Data Fig. 6e) and synchronicity (Fig. 2e) of TM-mediated ICWs in astrocytoma cells *in vivo*, but not in co-registered brain astrocytes of that tumour region. Inhibition of inositol triphosphate, which is gap-junction-permeable and the effector of gap-junction-mediated ICW propagation[26] also reduced ICWs between astrocytoma cells (Extended Data Fig. 6f).

The functional connection of single astrocytoma cells via TM-associated gap junctions was verified by rapid distribution of the gap junction-permeable dye sulforhodamine 101 (ref. 28) in the TM-connected cellular tumour cell network *in vivo* after local injection (Fig. 2f), which was inhibited by gap junction blockade (Extended Data Fig. 5g). Further experiments confirmed that another gap junction-permeable molecule was transferred between TM-connected cells (Extended Data Fig. 6h–j), while gap junction-impermeable large molecules were not (Extended Data Fig. 4c).

Connexin 43 connects TMs

To identify which of the known gap-junction-forming human connexins is involved in TM-mediated cell-to-cell communication in astrocytomas, we hypothesized that the deficiency of 1p/19q codeleted gliomas for intercellular TMs might also result in lower expression of the relevant TM-associated connexin(s). Analysis of 250 glioma samples of the TCGA data set revealed a list of differentially expressed genes between 1p/19q codeleted versus non-codeleted human tumours. Of the 20 connexins for which reads were mapped, only connexin 43 (also known as Cx43, GJA1) was differentially expressed, and found to be among the top 100 upregulated genes in 1p/19q non-codeleted tumours, both in *IDH* mutated and wild-type ones (Supplementary Table 1). This was confirmed in patient tumour tissue (Fig. 3a), and also in the primary glioma cell lines (Extended Data Fig. 6k). Confocal microscopy revealed punctate Cx43 immunoreactivity particularly at the TMs of astrocytoma cells (Fig. 3b), which was not seen for other connexins (Extended Data Fig. 6l). Remarkably, Cx43 immunoreactivity frequently located at the place where two different TMs crossed each

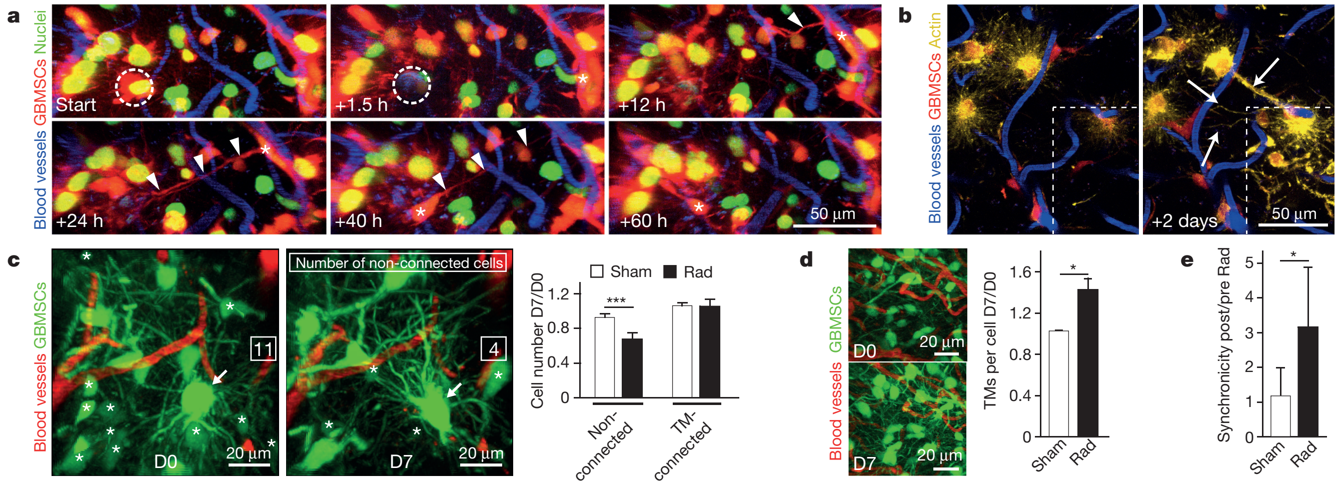


Figure 4 | TM-connected astrocytoma cell networks can repair themselves, and resist radiotherapy. a, Time series of exemplary 3D images (54 μm thickness) of a tumour region after laser-induced killing of a GMBSC (circle). Over time a TM (arrowheads) is extended, and a nucleus (asterisk) translocates via that TM to the place the killed cell had been located. **b**, Time course after laser-induced photodamage (dotted area), arrows: GBMSC TMs extending into the photodamaged region. MIP of 21 μm, $n = 3$ mice. **c**, GBMSC tumour microregions (3D images, 20 μm depth) before start of radiation (D0), and 7 days later (D7). Asterisks, non-TM-connected GBMSCs; arrow, one exemplary cell with many TMs. Graph, relative change of cellular subtypes under sham radiation (Sham), or treated with radiation at 3 × 7 Gy (Rad) ($n = 3$ mice per group, t-tests). **d**, Number of TMs per cell in these tumours (3D images left side, $n = 50$ cells per time point in $n = 3$ mice per group, Mann–Whitney test). **e**, Synchronicity of calcium transients after sham treatment or radiotherapy in individual GBMSC tumour regions (Fluo-4AM, $n = 3$ mice per group, Mann–Whitney test). All images, *in vivo* MPLSM. GBMSCs, S24 line. Error bars show s.d. $^{*}P < 0.05$, $^{***}P < 0.001$.

other (Fig. 3b). Contact sites of individual TMs with direct membrane–membrane contact were also detected with 3D SEM (Fig. 3c). Further analysis of our ICW data sets revealed that calcium waves can propagate via those crossings from one TM to another (Fig. 2a, red arrow).

To investigate the functional role of the Cx43 gap junction protein in astrocytoma progression, a stable short hairpin RNA knockdown of Cx43 was performed in GBMSCs. This significantly reduced the synchronicity of ICWs *in vivo* (Fig. 3d), and also the proportion of astrocytoma cells with multiple TMs at later time points (Extended Data Fig. 6m), which suggests a role for Cx43 gap-junction-mediated communication in long-term stabilization of TMs. In accordance with the proposed role of functional TMs for tumour progression, Cx43 deficiency resulted in reduced tumour size as observed by MRI (Fig. 3e) and improved survival (Extended Data Fig. 6n).

A self-repairing and resistant network

To investigate the role of TMs in damage repair *in vivo*, selective ablation of single GBMSCs was performed by applying a fatal laser dose to a fraction of their nuclear volume (1 μm^3). If the ablated cell was a prior member of the TM-connected network, new TMs were extended towards the dead cell, and within a few days a new nucleus advanced via those TMs to the location of the prior cell (Fig. 4a; $n = 8$ reconstitution events in 8 photodamaged tumour cells from $n = 3$ animals). If a non-TM-connected GBMSC was ablated, such repair mechanism was only infrequently observed (2/8 events in $n = 3$ animals; $P < 0.01$, Fisher's exact test). Photon damage to a larger volume consisting of 6–10 GBMSCs and normal brain parenchyma resulted in rapid extension of TMs of neighbouring GBMSCs into this area, followed by a marked increase in tumour cell density specifically in the damaged volume (Fig. 4b).

Next we investigated whether TM-connected tumour cell networks were also resistant against the cytotoxic effects of radiation therapy, a standard treatment of gliomas. While TM-connected cells were largely protected from cell death, unconnected tumour cells died in relevant numbers after radiotherapy (Fig. 4c, Extended Data Fig. 7a, b). Furthermore, TM-connected astrocytoma cells increased both their TM number (Fig. 4d) and their calcium communication (Fig. 4e, Extended Data Fig. 7c) as a reaction to radiotherapy. Concordantly, Cx43 knockdown reduced the radioprotective effect of TM interconnections, while non-TM-connected astrocytoma cells regressed like in control tumours (Extended Data Fig. 7d).

To explore potential mechanisms of TM-mediated protection from cytotoxicity, we measured basal intracellular calcium levels in astrocytoma cells before and during radiotherapy, using a ratiometric calcium indicator. Basal calcium levels were very homogeneous in non-irradiated cells, and also in TM-connected cells during radiotherapy, while unconnected cells developed a high variability of their intracellular calcium levels during irradiation (Extended Data Fig. 7e–h).

Drivers of TM formation

Next we sought to identify the crucial molecular pathways that drive the formation of TMs to better understand their nature, and to substantiate their role for tumour progression and resistance. For this purpose, we first analysed the *in silico* data set of 1p/19q non-codeleted versus codeleted human gliomas (Supplementary Table 1) by using Ingenuity Pathway Analysis. Here, biological functions that were prominently activated in 1p/19q non-codeleted astrocytomas included "cellular movement" and "cell-to-cell signalling and interaction", supporting the proposed function of TMs in these tumours (Extended Data Fig. 8a, b for *IDH* mutant astrocytomas; similar results for *IDH* wild-type astrocytomas, data not shown). Intriguingly, we found many canonical pathways involved in the outgrowth of neurites, and neurite-like membrane protrusions to be more activated in 1p/19q non-codeleted gliomas, including integrin[29], phospholipase C[30], Rho family GTPases[31], HMGB1[32], and also the prototypical neurotrophin/TRK signalling pathways[30] (Extended Data Fig. 8c; confirmed in *IDH* wild-type astrocytomas, data not shown). The latter was confirmed at the protein level in human gliomas, where the neurotrophins NGF (located on 1p) and NT-4 (19q) were downregulated in 1p/19q codeleted tumours, and also their respective membrane receptors TrkA and TrkB, which has been described before[33] (Extended Data Fig. 9a).

When considering these results and reviewing the literature for known downstream effectors particularly relevant for the formation of neurite-like membrane protrusions, the growth-associated protein GAP-43 came into focus. GAP-43 is highly expressed in axonal growth cones[34,35], induced by neurotrophin receptor signalling[36,37], and drives neuronal progenitor cell migration[38]. Remarkably, GAP-43

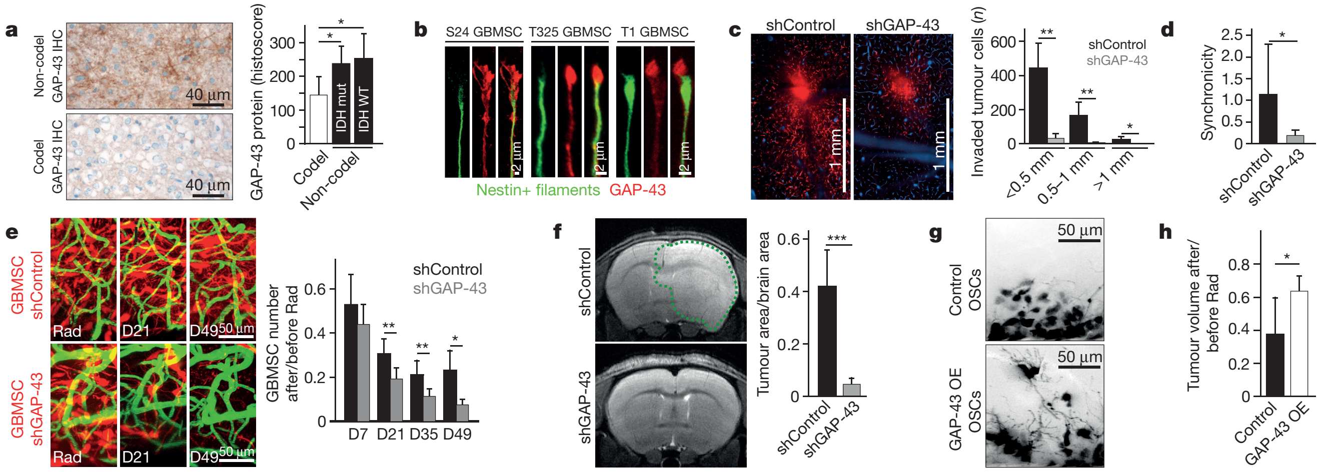


Figure 5 | GAP-43 is required for TM outgrowth and function. **a**, GAP-43 protein expression (immunohistochemistry) in 1p/19 non-codeleted (*IDH* wild-type and *IDH* mutated) versus codeleted gliomas ($n = 8$ each) (ANOVA on the ranks, Student–Newman–Keuls post hoc test). **b**, Immunocytological images of GAP-43 protein, with preferential GAP-43 localization at the tip of TMs. **c**, Single-plane images of control and shGAP-43 knockdown GBMSC S24 tumours; right, corresponding quantification ($n = 3$ mice per condition; Mann–Whitney tests). **d**, Synchronicity analysis of calcium peaks in S24 shControl versus shGAP-43 cells *in vivo* ($n = 3$ mice; Mann–Whitney test). **e**, Time course after irradiation of S24 shControl versus shGAP-43 tumours, corresponding quantifications ($n = 3$ mice per group; *t*-tests). **f**, MRI images of S24 shControl versus shGAP-43 tumours, 60 days after radiation (115 days after tumour implantation); right, quantifications of 5–6 animals per group (*t*-test). **g**, GAP-43 overexpression in BT088 oligodendroglial stem-like cell lines (OSCs) versus controls, 14 days after injection. **h**, Relative change of tumour volumes 21 days after radiotherapy in BT088 vector-control versus GAP-43 overexpression tumours ($n = 3$ mice per group; *t*-test). **c**–**e**, **g**, **h**, *In vivo* MPLSM. Error bars show s.d. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

overexpression is sufficient for the outgrowth of membrane tubes in neuronal[39] and even in non-neuronal[40] cells. Indeed, GAP-43 was significantly higher expressed in 1p/19q non-codeleted human gliomas (Fig. 5a) and primary stem-like cell lines (Extended Data Fig. 9b, c) when compared with 1p/19q codeleted ones. Of note, GAP-43 preferentially localized to the cone-like and nestin-negative tips of sprouting TMs (Fig. 5b), similar to its known enrichment in the nerve growth cone[35].

To interfere with TM formation during astrocytoma progression, we engineered GBMSCs with a genetic knockdown of GAP-43. While *in vitro* viability of these cells was not affected (data not shown), their TMs *in vivo* were structurally abnormal, with reduced branching (Extended Data Fig. 9d). This was associated with an impaired dissemination of tumour cells (Fig. 5c, Extended Data Fig. 9e), resulting from both their decreased mean invasion speed (Extended Data Fig. 9f) and proliferation capacity (Extended Data Fig. 9g) in the mouse brain. Importantly, intercellular TM connections (Extended Data Fig. 9h) and ICWs (Fig. 5d) were reduced in GAP-43 deficient tumours, which was accompanied by a selective reduction of Cx43 gap junction protein expression (Extended Data Fig. 9i). These deficiencies in TM-mediated features of tumour progression lead to a marked reduction of tumour size in the mouse brain (Extended Data Fig. 9j), and to an improved survival of the animals (Extended Data Fig. 9k). When radiotherapy was applied, GAP-43 deficiency resulted in an increased regression of tumour cells, as revealed by repetitive *in vivo* MPLSM (Fig. 5e). This was confirmed by MRIs 60 days after radiation, where no relevant tumour-derived signal changes were detectable in shGap43 tumour bearing animals, while control tumours were large, causing neurological symptoms in mice (Fig. 5f). Histological analysis at this time point confirmed only small remnants of proliferation-deficient tumour cells in GAP-43 knockdown tumours (Extended Data Fig. 9l).

Finally, we overexpressed the GAP-43 protein in 1p/19q codeleted primary oligodendroglioma cells, to achieve protein levels comparable to the 1p/19q non-codeleted GBMSC lines (Extended Data Fig. 9m). This lead to a morphological shift to a TM-rich, thus astrocytoma-like phenotype of GAP-43 overexpressing oligodendroglioma cells (Fig. 5g, Extended Data Fig. 9n, o). Remarkably, the induction of TM formation in these tumours resulted in an increase in tumour cell invasion into the brain (Fig. 5g, Extended Data Fig. 9p), and also in an increase in radioresistance (Fig. 5h); both were comparably low in control oligodendrogliomas.

Conclusions

The ability to interconnect via ultra-long and highly functional TMs is an important mechanism of progression and resistance in astrocytomas, and depends on molecular pathways that are active when 1p/19q is intact. The resulting multicellular network is able to communicate via Cx43 gap junction connections. Multiple functions have been reported previously for different connexins in glioma pathology[41,42]. This includes Cx43, which is also highly expressed in non-malignant astrocytes, connecting them to functional and resistant[43] cellular networks. Here we provide one possible mechanism for this resistance: maintenance of calcium homeostasis by network integration. Since increases of intracellular calcium levels are required for radiotherapy-induced cytotoxicity[44], and even small calcium increases are involved in intrinsic apoptotic cell death in glioma cells[45], one can assume that intercellular TMs can serve as a means for an individual tumour cell to distribute critical elevations of small molecules like calcium within the larger network, achieving nonlethal levels.

The data presented here support the notion that tumours are complex organs, which has so far been attributed to the supportive contribution of non-malignant cell types[46], including neurons in brain tumours[47]. Our study adds to this concept by demonstrating that single cancer cells within one tumour communicate and cooperate with each other in a complex but ordered manner that is by itself reminiscent of a functional organ. It has become clear that tumours can hijack programs that are part of normal tissue development[48]. The key finding of this study is that TMs, which are generated by GAP-43 just as axons are in neurons, allow efficient tumour progression, network communication, and resistance to adverse events (Extended Data Fig. 10). Thus we anticipate that pharmacological targeting of TM formation and function will open new therapeutical avenues for treatment-resistant brain tumours.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 December 2014; accepted 9 October 2015.
Published online 4 November; corrected online 2 December 2015 (see full-text HTML version for details).

1. Cuddapah, V. A., Robel, S., Watkins, S. & Sontheimer, H. A neurocentric perspective on glioma invasion. *Nature Rev. Neurosci.* **15,** 455–465 (2014).
2. The Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* **372,** 2481–2498 (2015).
3. Eckel-Passow, J. E. *et al.* Glioma groups based on 1p/19q, *IDH*, and *TERT* promoter mutations in tumors. *N. Engl. J. Med.* **372,** 2499–2508 (2015).
4. Wiestler, B. *et al.* Integrated DNA methylation and copy-number profiling identify three clinically and biologically relevant groups of anaplastic glioma. *Acta Neuropathol.* **128,** 561–571 (2014).
5. van den Bent, M. J. *et al.* Adjuvant procarbazine, lomustine, and vincristine chemotherapy in newly diagnosed anaplastic oligodendroglioma: long-term follow-up of EORTC brain tumor group study 26951. *J. Clin. Oncol.* **31,** 344–350 (2013).
6. Cairncross, G. *et al.* Phase III trial of chemoradiotherapy for anaplastic oligodendroglioma: long-term results of RTOG 9402. *J. Clin. Oncol.* **31,** 337–343 (2013).
7. Ramírez-Weber, F. A. & Kornberg, T. B. Cytonemes: cellular processes that project to the principal signaling center in *Drosophila* imaginal discs. *Cell* **97,** 599–607 (1999).
8. Rustom, A., Saffrich, R., Markovic, I., Walther, P. & Gerdes, H. H. Nanotubular highways for intercellular organelle transport. *Science* **303,** 1007–1010 (2004).
9. Lou, E. *et al.* Tunneling nanotubes provide a unique conduit for intercellular transfer of cellular contents in human malignant pleural mesothelioma. *PLoS One* **7,** e33093 (2012).
10. Sowinski, S. *et al.* Membrane nanotubes physically connect T cells over long distances presenting a novel route for HIV-1 transmission. *Nature Cell Biol.* **10,** 211–219 (2008).
11. Inaba, M., Buszczak, M. & Yamashita, Y. M. Nanotubes mediate niche-stem-cell signalling in the *Drosophila* testis. *Nature* **523,** 329–332 (2015).
12. Hsiung, F., Ramirez-Weber, F. A., Iwaki, D. D. & Kornberg, T. B. Dependence of *Drosophila* wing imaginal disc cytonemes on Decapentaplegic. *Nature* **437,** 560–563 (2005).
13. Smith, I. F., Shuai, J. & Parker, I. Active generation and propagation of Ca^{2+} signals within tunneling membrane nanotubes. *Biophys. J.* **100,** L37–L39 (2011).
14. Wang, X., Veruki, M. L., Bukoreshtliev, N. V., Hartveit, E. & Gerdes, H. H. Animal cells connected by nanotubes can be electrically coupled through interposed gap-junction channels. *Proc. Natl Acad. Sci. USA* **107,** 17194–17199 (2010).
15. Sherer, N. M. Long-distance relationships: do membrane nanotubes regulate cell-cell communication and disease progression? *Mol. Biol. Cell* **24,** 1095–1098 (2013).
16. Kienast, Y. *et al.* Real-time imaging reveals the single steps of brain metastasis formation. *Nature Med.* **16,** 116–122 (2010).
17. Lee, J. *et al.* Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell* **9,** 391–403 (2006).
18. Lowery, L. A. & Van Vactor, D. The trip of the tip: understanding the growth cone machinery. *Nature Rev. Mol. Cell Biol.* **10,** 332–343 (2009).
19. Friedl, P. & Alexander, S. Cancer invasion and the microenvironment: plasticity and reciprocity. *Cell* **147,** 992–1009 (2011).
20. Scherer, H. J. The forms of growth in gliomas and their practical significance. *Brain* **63,** 1–35 (1940).
21. Zhu, Z. *et al.* Targeting self-renewal in high-grade brain tumors leads to loss of brain tumor stem cells and prolonged survival. *Cell Stem Cell* **15,** 185–198 (2014).
22. Calabrese, C. *et al.* A perivascular niche for brain tumor stem cells. *Cancer Cell* **11,** 69–82 (2007).
23. Capper, D., Zentgraf, H., Balss, J., Hartmann, C. & von Deimling, A. Monoclonal antibody specific for *IDH1* R132H mutation. *Acta Neuropathol.* **118,** 599–601 (2009).
24. Kuchibhotla, K. V., Lattarulo, C. R., Hyman, B. T. & Bacskai, B. J. Synchronous hyperactivity and intercellular calcium waves in astrocytes in Alzheimer mice. *Science* **323,** 1211–1215 (2009).
25. Cornell-Bell, A. H., Finkbeiner, S. M., Cooper, M. S. & Smith, S. J. Glutamate induces calcium waves in cultured astrocytes: long-range glial signaling. *Science* **247,** 470–473 (1990).
26. Leybaert, L. & Sanderson, M. J. Intercellular Ca^{2+} waves: mechanisms and function. *Physiol. Rev.* **92,** 1359–1392 (2012).
27. Weissman, T. A., Riquelme, P. A., Ivic, L., Flint, A. C. & Kriegstein, A. R. Calcium waves propagate through radial glial cells and modulate proliferation in the developing neocortex. *Neuron* **43,** 647–661 (2004).
28. Nimmerjahn, A., Kirchhoff, F., Kerr, J. N. & Helmchen, F. Sulforhodamine 101 as a specific marker of astroglia in the neocortex *in vivo*. *Nature Methods* **1,** 31–37 (2004).
29. Tomaselli, K. J., Neugebauer, K. M., Bixby, J. L., Lilien, J. & Reichardt, L. F. N-cadherin and integrins: two receptor systems that mediate neuronal process outgrowth on astrocyte surfaces. *Neuron* **1,** 33–43 (1988).
30. Stephens, R. M. *et al.* Trk receptors use redundant signal transduction pathways involving SHC and PLC-γ1 to mediate NGF responses. *Neuron* **12,** 691–705 (1994).
31. Kozma, R., Sarner, S., Ahmed, S. & Lim, L. Rho family GTPases and neuronal growth cone remodelling: relationship between increased complexity induced by Cdc42Hs, Rac1, and acetylcholine and collapse induced by RhoA and lysophosphatidic acid. *Mol. Cell. Biol.* **17,** 1201–1211 (1997).
32. Fang, P., Schachner, M. & Shen, Y. Q. HMGB1 in development and diseases of the central nervous system. *Mol. Neurobiol.* **45,** 499–506 (2012).
33. Wang, Y. *et al.* Trk A, B, and C are commonly expressed in human astrocytes and astrocytic gliomas but not by human oligodendrocytes and oligodendroglioma. *Acta Neuropathol.* **96,** 357–364 (1998).
34. Goslin, K., Schreyer, D. J. & Skene, J. H. & Banker, G. Development of neuronal polarity: GAP-43 distinguishes axonal from dendritic growth cones. *Nature* **336,** 672–674 (1988).
35. Skene, J. H. *et al.* A protein induced during nerve growth (GAP-43) is a major component of growth-cone membranes. *Science* **233,** 783–786 (1986).
36. Lavenius, E., Gestblom, C., Johansson, I., Nånberg, E. & Påhlman, S. Transfection of TRK-A into human neuroblastoma cells restores their ability to differentiate in response to nerve growth factor. *Cell Growth Differ.* **6,** 727–736 (1995).
37. Koponen, E., Lakso, M. & Castrén, E. Overexpression of the full-length neurotrophin receptor trkB regulates the expression of plasticity-related genes in mouse brain. *Brain Res. Mol. Brain Res.* **130,** 81–94 (2004).
38. Haag, D. *et al.* Nos2 inactivation promotes the development of medulloblastoma in $Ptch1^{+/-}$ mice by deregulation of Gap43-dependent granule cell precursor migration. *PLoS Genet.* **8,** e1002572 (2012).
39. Aigner, L. *et al.* Overexpression of the neural growth-associated protein GAP-43 induces nerve sprouting in the adult nervous system of transgenic mice. *Cell* **83,** 269–278 (1995).
40. Zuber, M. X., Goodman, D. W., Karns, L. R. & Fishman, M. C. The neuronal growth-associated protein GAP-43 induces filopodia in non-neuronal cells. *Science* **244,** 1193–1195 (1989).
41. Sin, W. C., Crespin, S. & Mesnil, M. Opposing roles of connexin43 in glioma progression. *Biochim. Biophys. Acta* **1818,** 2058–2067 (2012).
42. Hitomi, M. *et al.* Differential connexin function enhances self-renewal in glioblastoma. *Cell Reports* **11,** 1031–1042 (2015).
43. Le, H. T. *et al.* Gap junction intercellular communication mediated by connexin43 in astrocytes is essential for their resistance to oxidative stress. *J. Biol. Chem.* **289,** 1345–1354 (2014).
44. Tombal, B., Denmeade, S. R., Gillis, J. M. & Isaacs, J. T. A supramicromolar elevation of intracellular free calcium ($[Ca^{2+}]_i$) is consistently required to induce the execution phase of apoptosis. *Cell Death Differ.* **9,** 561–573 (2002).
45. McFerrin, M. B., Turner, K. L., Cuddapah, V. A. & Sontheimer, H. Differential role of IK and BK potassium channels as mediators of intrinsic and extrinsic apoptotic cell death. *Am. J. Physiol. Cell Physiol.* **303,** C1070–C1078 (2012).
46. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144,** 646–674 (2011).
47. Venkatesh, H. S. *et al.* Neuronal activity promotes glioma growth through neuroligin-3 secretion. *Cell* **161,** 803–816 (2015).
48. Egeblad, M., Nakasone, E. S. & Werb, Z. Tumors as organs: complex tissues that interface with the entire organism. *Dev. Cell* **18,** 884–901 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Ruiz de Almodovar and H.-H. Gerdes for discussions and comments; P. Rübman, B. Kast, A. Habel, A. Tietz-Dalfuβ and M. Fischer for technical assistance; R. Hermann for help with vibratome slices; G. Eisele for providing the WJ cell line; P. Friedl for the Lifeact-YFP-construct and the $IDH1^{R132H}$ thick section staining protocol; H. Glimm for the pCCL.PPT. SFFV.MCS.IRES.eGFP.WPRE-vector backbone; and M. Splinter, M. Brand, C. Lang for help with radiation experiments. This work was funded by grants from the German Research Foundation (DFG, WI 1930/5-1 (F.W.) and Major Equipment Grant INST 114089/26-1 FUGG (F.W., W.W.)), an intramural grant from the DKFZ to F.W. and H.L., Heinrich F. C. Behr-Stipend to S. Weil. F.S. is a fellow of the Medical Faculty Heidelberg PostDoc-Program. The results published here are in part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

Author Contributions F.W., M.O. and W.W. were responsible for experimental design, data interpretation, and writing of the manuscript. M.O., E.J., S. Weil and Y.L. performed MPLSM experiments. F.S. and A.v.D. performed stainings and analyses of human glioma tissues. M.O., M.G., E.J., S. Weil performed cell culture and cranial window implantations. G.S. was responsible for quantification and analysis of the calcium data. T.K., H.H., V.V. provided electron microscopy data and corresponding analyses. B.W. performed the TCGA data analysis. F.T.K. and S.H. collected MRI data. J.B. and T.S., M.R. and K.K.J. performed cell culture experiments, S.P. and D.L. established and characterized cell lines. A.S.B., L.H. and M. Preusser conducted histological experiments. V.H. and C.S. constructed the rrl-CAG-lGC3 vector. O. Griesbeck, G.K. and O. Garaschuk constructed the Twitch-3 vector, and interpreted the calcium imaging data. M.S. performed analyses of thick human tumour slices. M. Pauli conducted electroporation experiments. P.H. and P.E.H. were responsible for radiation. D.K. performed analysis of image data and confocal image acquisition. M. Platten performed data interpretation. M.J. performed FACS sorting. H.M. and S. Weiss provided cell lines and interpreted data. H.L. provided the syngeneic tumour model.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.W. (frank.winkler@med.uni-heidelberg.de).

METHODS

Animals, and surgical procedures. 8–10 weeks old male NMRI nude mice were used for all studies with human primary brain tumour cells. As a syngeneic astrocytoma mouse model we used Nestin-Tv-a;Tlx-GFP mice in combination with RCAS-PDGFB/AKT vectors[21]. All animal procedures were performed in accordance with the institutional laboratory animal research guidelines after approval of the Regierungspräsidium Karlsruhe, Germany (governmental authority). All efforts were made to minimize animal suffering and to reduce the number of animals used. Mice were clinically scored and if they showed marked neurological symptoms or weight loss of >20%, experiments were terminated. In none of the experiments these limits were exceeded. No maximum tumour size was defined for the invasive brain tumour models.

Cranial window implantation in mice was done in a modification of what has been previously described[49], including a custom-made titanium ring for painless head fixation during imaging.

2–3 weeks after cranial window implantation, 30,000 tumour cells were stereotactically injected into the mouse brain at a depth of 500 µm. For survival experiments we injected 50,000 tumour cells. In a subgroup of mice, a short plastic tube was glued under the glass, with one end inside and one outside, which allowed topical application of different substances under the window without the need to re-open it.

For intratumoral microinjection of sulforhodamine 101 (SR101, Molecular Probes, S-359), 50 nl of 100 µM SR101 (with or without 100 µM CBX, Sigma-Aldrich, C4790) were injected with a very thin glass pipette into tumour regions of similar cellular densities, >90 days after tumour injection.

Radiation treatment. Established tumours were irradiated with 7 Gy on three consecutive days (total dose 21 Gy) in regions matching in tumour cell density using a 6 MV linear accelerator with a 6 mm collimator (adjusted to the window size) at a dose rate of 3 Gy min^{-1} (Artiste, Siemens), or no radiation was applied (sham radiation), at day 60 (±10 days) after tumour implantation. For MRI studies, a total brain radiation with the same dose and a field size of 17 mm × 250 mm (allowing the irradiation of several mice) was used. The used radiation schedule is in the range of the commonly prescribed 60 Gy in 2 Gy fractions for malignant glioma patients, assuming an α/β of ~ 10 in the linear quadratic model and taking into account the radiation time of 3 days.

***In vivo* multiphoton laser scanning microscopy (MPLSM).** MPLSM imaging was done with a Zeiss 7MP microscope (Zeiss) equipped with a Coherent Chameleon UltraII laser (Coherent). The following wavelengths were used for excitation: 750 nm (dsRed, FITC-dextrane, tdTomato), 840 nm (Fluo-4AM), 850 nm (GFP, TRITC-dextrane, Rhod-2AM), 860 nm (CFP, for Förster Resonance Energy Transfer (FRET) imaging) and 950 nm (tdTomato, YFP). Appropriate filter sets (band pass 500–550 nm/band pass 575–610 nm and band pass 460–500 nm/band pass 525–560 nm (for FRET)) were used. Standard settings for imaging were gains between 650 and 750 nm (depending on the depth, the fluorescence intensity of the fluorophore and the window quality), and a *z*-interval of 3 µm. Laser power was tuned as low as possible.

The body temperature of mice was kept constant using a rectal thermometer and a heating pad. Isoflurane concentration (in 100% O_2) was chosen as low as possible (0.5–1.5%) to avoid interference with the calcium communication between astrocytoma cells. Fluorescent dextranes (FITC (2M MW)- or TRITC (500.000 MW)-conjugated, 10 mg ml^{-1}, Sigma) were injected intravenously to obtain angiograms.

For *in vivo* ablation of single astrocytoma cells, only the volume of the GFP-labelled cell nucleus was exposed to continuous scanning with a high power laser beam until disintegration of the nucleus became visible. To investigate the reaction of TMs after the photodamage of a wider brain region, a larger volume ($0.5–1 \times 10^6$ μm^3) was scanned repetitively for approximately 8 min with high power, resulting in a total photon dose that was >50 times higher than during "diagnostic" imaging.

***In vivo* calcium imaging with MPLSM.** The following small molecule calcium indicators were applied to the brain surface for 45 min: for GFP-transfected tumour cells, 2 mM Rhod-2AM (Life Technologies, R-1244); for RFP-transfected, 2 mM Fluo-4AM (Life Technologies, F-14201). Pharmacological gap junction inhibition was achieved by superfusion with the inhibitor CBX (100 µM; control substance: extracellular saline; $n = 3$ mice per group). Other superfused substances were suramin (100 µM, ATP antagonist) and 2-aminoethoxydiphenyl borate (2-APB, 100 µM, inhibitor of inositol triphosphate receptors). Two genetically encoded calcium indicators were lentivirally transduced to GBMSCs: the Lck-GCaMP3 sensor in the rrl-CAG-lGC3 vector (CAG promoter to control expression of DsRed and the Ca^{2+} sensor that monitors near-membrane changes in $[Ca^{2+}]_i$[50]). The ratiometric calcium sensor Twitch-3 was used to determine intracellular calcium concentrations by FRET as previously described[51].

MRI studies. MRI images were obtained at day 72 after tumour implantation for non–irradiated animals, and at day 115 for irradiated mice (60 days after radiotherapy; time points were chosen when first control animals developed neurological symptoms and/or lost 20% weight, and had to be euthanized). All scans were performed on a 9.4 T horizontal bore MR scanner (BioSpec 94/20 USR, Bruker BioSpin GmbH) with a four channel phased array surface coil. A T2-weighted rapid acquisition with refocused echoes (RARE) sequence was acquired to determine tumour volume.

Cell lines and cell culture experiments. Tumour cell lines derived from resected glioblastomas were cultivated in DMEM-F12 under serum-free non-adherent, 'stem-like' conditions, including B27 supplement (12587-010, Gibco), insulin, heparin, epidermal growth factor and fibroblast growth factor[17,52] (GBMSCs: P3, S24, T1, T269, T325, WJ). These 6 GBMSC lines were selected because they were capable of growing to tumours in mouse brains; all were non-codeleted for 1p/19q, and *IDH* wild-type. Two oligodendroglioma cell lines harbouring the typical 1p/19q codeletion (BT088 and BT054) were kept under the same cell culture conditions[53]. Of note, BT054 is $IDH1^{R132H}$ mutated, while BT088 has lost the heterozygous $IDH1^{R132H}$ mutation of the patient tumour it was derived from, but still maintains its GCIMP phenotype (data not shown). Typical genetic changes of glioblastoma were confirmed for S24 using comparative genomic hybridization (CGH, see Extended Data Fig. 1i); the T1, T269, T325 and WJ lines had been characterized before[52], as well as the P3 line[54]. Cells were regularly checked for mycoplasma infections and authenticity (species control).

Tumour cells were transduced with lentiviral vectors for multicolour imaging. For cytosolic GFP expression, we used the pLKO.1-puro-CMV-TurboGFP_shnon-target-vector (SHC016 Sigma Aldrich), for cytosolic RFP (tdTomato) expression the LeGo-T2 vector (gift from A. Trumpp), and nuclear GFP expression (H2B–GFP) was achieved by transduction with pLKO.1-LV-GFP (Addgene 25999, Elaine Fuchs). Transduction with pLenti6.2 hygro/V5-Lifeact-YFP made it possible to image the *in vivo* dynamics of actin filaments, FUmGW (Addgene 22479, Connie Cepko) allowed *in vivo* illustration of cell membranes. Microtubuli were marked using the LentiBrite GFP-Tubulin Lentiviral Biosensor (17-10206, Merck Millipore). Lentiviral particles were produced as described before[55]. For *in vivo* tracking of Myosin II, a plasmid transfection with FuGENE HD (Promega) was performed with the Myosin-IIA-GFP vector (Addgene 38297, Matthew Krummel).

Production of lentiviral knockdowns of Cx43 (pLKO1.1-puro-CMV-tGFP-vector, Sigma Aldrich, target sequence: GCCCAAACTGATGGTGTCAAT) and GAP-43 (pLKO1.1-puro-CMV-vector, Sigma Aldrich, target sequence: TGTAGATGAAACCAAACCTAA) by shRNA technology was carried out as described before[55]. Control cells were infected with the appropriate non_target shRNA-lentiviral particles (SHC016, Sigma Aldrich). For overexpression of GAP-43, the open reading frame of GAP-43 was cloned into the pCCL.PPT.SFFV.MCS.IRES.eGFP.WPRE-vector backbone. Lentiviral particle production and transduction of target cells was done as described before[55].

Tumour cells were incubated with the harvested virus and 8 mg ml^{-1} polybrene (Merck Millipore) for 24 h. Quantification revealed a 80% protein knockdown for Cx43 and a 92.5% for GAP-43 (Western Blot analyses). If necessary, tumour cells were selected for the fluorophores by FACS sorting (BD FACSAria II Cell Sorter) or antibiotics.

For tracking of mitochondria, the BacMam 2.0 technology was used (CellLight Mitochondria-GFP, BacMam 2.0, C10600, Life Technologies).

Immunohistochemistry (IHC) and immunocytochemistry (ICC). For IHCs and ICCs, standard protocols were used. For human brain analyses, thin (3 µm) formalin-fixed paraffin-embedded human tissue sections from resected primary gliomas were obtained from the Department of Neuropathology in Heidelberg in accordance with local ethical approval. Human sections were incubated with anti-BRAF-V600E (VE1, Ventana), anti-IDH1 R132H (H09, Dianova), anti-Cx43 (C6219, Sigma), anti-GAP-43 (8945, Cell Signaling), anti-NGF (ab52918, Abcam), anti-NT4 (ab150437, Abcam), anti-TrkA (ab76291, Abcam) and anti-TrkB (ab134155, Abcam) antibodies. If not explicitly stated, all oligodendrogliomas had a 1p/19q codeletion, and all astrocytomas were non-codeleted for 1p/19q. To detect contralateral tumour cells in human brains, large sections were analysed as previously described[56].

For mouse brain analyses, animals were transcardially perfused with PBS followed by 4.5% paraformaldehyde (PFA). For ICCs, cells were grown on glass slides for 4 days and fixed with PFA. The following antibodies were used for 10 µm cryotome sections and ICCs: anti-nestin (ab6320, Abcam, specific staining of GBMSCs, no signal detectable in normal mouse brain), in combination with anti-β-catenin (ab16051, Abcam), anti-β-parvin (sc-50775, Santa Cruz), anti-beta tubulin III (ab18207, Abcam), anti-Cx26 (ab59020, Abcam), anti-Cx31 (ab156582, Abcam), anti-Cx37 (ab185820, Abcam), anti-Cx43, anti-GAP-43, anti-GFAP (Z0334, Dako), anti-Ki-67 (M7240, Dako), anti-myosin IIa (ab24762, Abcam), anti-myosin X (22430002, Novus Bioscience), anti-N-cadherin (ab18203, Abcam), and anti-PDI (ab3672, Abcam).

Photo-oxidation of GFP, and serial section scanning electron microscopy (SEM). Ten S24 GBMSC brain tumour tissue blocks were prepared for photo-oxidation as described previously[57]. Serial 70 nm thick sections of photo-oxidated epon-embedded tissue were produced using 3D SEM as described before[58]. A volume of 747 μm^3 was imaged. Specimens were imaged with a Zeiss 1530 scanning electron microscope. Images were aligned manually.

Western Blot. Western blots were performed according to standard protocols. Total protein lysates (20–50 μg) were electrophoretically separated using a 10% SDS–PAGE. After blotting and blocking, the primary antibodies (see above) were incubated over night at 4 °C. As loading control, anti-GAPDH antibody (C4780, Linaris), or anti-alpha-tubulin antibody (T9026, Sigma) was used.

Electroporation/microinjection. Horizontal acute brain slices were obtained from 2 NMRI nude mice with 131 days old S24 as described[59]. Patch electrodes with resistances of 5–10 MΩ were filled with Lucifer yellow (5 mg ml^{-1}, L0259, Sigma) and approached to identified tumour cells under visual control using a 63×, NA 1.0 dipping lens (Zeiss). The dye was transferred into tumour cells with an Axoporator 800A (Axon instruments) by 1-ms square voltage pulses at 50 Hz. Pulse amplitude was adjusted between −5 V and −20 V and train duration was adjusted up to 3 s to receive sufficient labelling of the target cell.

Image processing. MPLSM images were acquired by the ZEISS ZEN Software (Zeiss, Germany). After primary image calculation (for example, subtraction of different channels to remove unspecific background), images were transferred to Imaris (Bitplane, Switzerland) to allow 3D visualization, rendering and analysis of the data. For illustration of different aspects single planes, maximum intensity projections (MIPs) or 3D images were used. For exemplary illustration of tumour cell interconnectivity and TM branches, *z*-stacks were rendered manually (tumour cell bodies with surface function; TMs with filament tracker function). When a TM started at one cell and ended at another, these cells were defined as connected. Serial electron micrographs were reconstructed using OpenCAR software. 3D analysis of electron microscopy images was done using the Amira 5.4.6 software (Visage Imaging, Richmond, Australia). Some of the data (for example, calcium imaging) were transferred to the ImageJ software (Rasband, W.S., ImageJ, NIH). Videos were extracted from ZEN or Imaris and edited in Adobe Premiere Pro CS6.

Quantification of histology and MPLSM imaging data. In patient tumour tissue (only from primary tumour resections), maximum TM length was measured in standard 3 μm thin *IDH1*^{R132H} IHC sections. Here, TMs were divided into 3 groups: <50 μm (not qualifying as definite TM, because other cellular structures might still be confused with filamentous structures of this length); shorter TMs of 50–100 μm, and longer TMs of >100 μm length. Quantitative analysis of human IHCs was done by a Histoscore (range 0–300) as described before[60,61].

For *in vivo* imaging data, TM numbers, branches per TM and connections per cell were counted manually, and TM lengths were measured manually in the slice mode in Imaris. Cells without a TM connection were defined as "non-connected" and cells with at least one TM-connection as "connected". TMs were also classified as connecting when the connected cell was outside the region of interest. To analyse the number of TMs before and after irradiation, the TMs of individual, identical cells were counted at both time points. The mean speed of tumour cell invasion in S24 shControl versus shGAP-43 tumours was determined by analysis of three consecutive imaging time points within a 24 h time interval *in vivo*. Distances of tumour cells to the main tumour mass (defined as a radius of 0.5 mm around the middle of the main tumour) were analysed and grouped, or displayed as individual distances to the main tumour core in tumours that were much less invasive (oligodendrogliomas). Nuclei and mitochondria (time-lapse imaging data) were marked using the spot function of Imaris. They were connected to tracks and the mean track speed was calculated. For quantification and analysis of fluorescence intensity after SR101 application, all GFP-expressing tumour cells in a volume were marked using the spot function of Imaris and then mean intensities of the SR101-channel of these spots were calculated and compared with each other. For quantification of tumour volumes two regions per animal were marked using the surface function of Imaris.

Quantification and analysis of calcium imaging data. Tumour cells and non-malignant brain astrocytes were identified by GFP/RFP expression and uptake of the chemical calcium indicator, and marked manually by the use of the region of interest manager of ImageJ. Mean grey values were measured over time. This data was processed by the program GNU Octave 3.8.1 (John W. Eaton, GPL): images were normalized to the background fluorescence using a sliding interval of ±10 images. Local maxima of calcium signals were detected by the findpeaks function (signal package, Octave-Forge). Thus the number of calcium peaks of each cell (N) could be determined and the frequency (f) was calculated. The frequency was standardized for the cell number of each region. Synchronous cells, the number of synchronous communications, and the time point of the synchronous firing were determined. Analysis was done in a window of 2 frames around each peak. This allowed to assess the synchronicity S ($S \in \mathbb{R}^+ \cup \{0\}$), which was defined by us as the fraction of the whole number of synchronous cells (N_{Syn}) divided by the number of calcium peaks for the given cell (N_{Ca}). In case the cells were not active, a synchronicity of zero was allotted.

$$\text{Synchronicity } S = \begin{cases} \frac{N_{\mathrm{Syn}}}{N_{\mathrm{Ca}}} & \text{for} \quad N_{\mathrm{Ca}} > 0 \\ 0 & \text{for} \quad N_{\mathrm{Ca}} = 0 \end{cases}$$

Hence, synchronicity states the average number of interactions at the same time point. For example, in a system with a synchronicity of 1, a firing cell interacts with a second one; for a synchronicity of 10, one cell is communicating with 10 other cells.

For the comparison between different blockers *in vivo*, the synchronicity was normalized to the baseline level. Finally, the results were summed up by a heat map. The number of calcium peaks of these cells were coded by a colour map. Synchronic cells were connected by lines, whereat the colour described the time point of the synchronic firing.

For measurement of relative changes in fluorescence intensity, tumour cells were again marked manually and relative changes were calculated ($\Delta F/F_0$). F_0 was defined as the average intensity of the 20% lowest grey values in a region of interest.

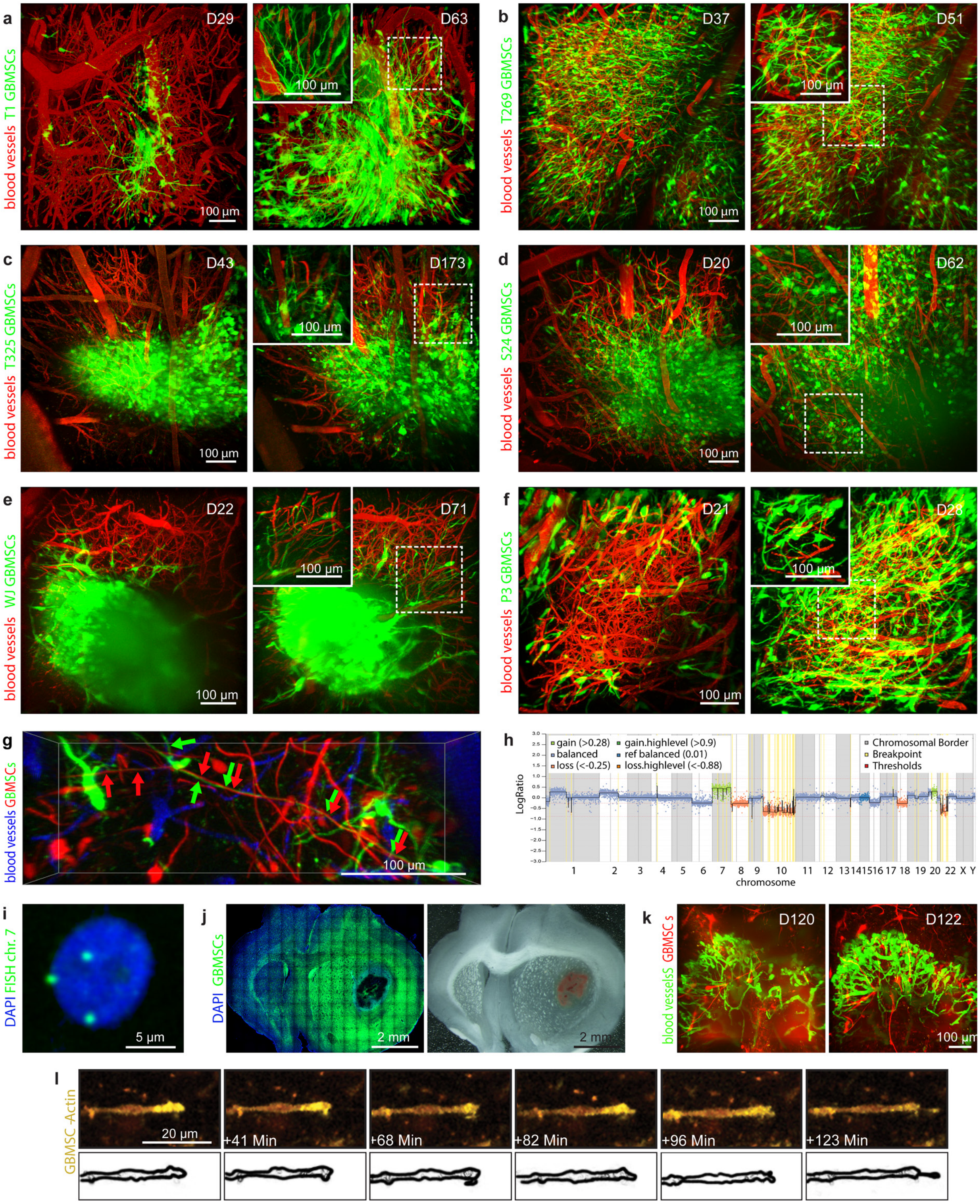
Quantification and analysis of MRI images. The slice with the largest tumour area per mouse was chosen, and both the tumour (hyperintense on T2-weighted images) and the whole brain were segmented manually. The ratio of these two areas were determined and compared between the different groups ($n = 6$ mice per group, *t*-tests).

Functional characterization of differential mRNA expression of human gliomas. RNA sequencing raw data (mapped to genes) and curated IDH-1/2 mutation data were downloaded from The Cancer Genome Atlas (TCGA) data portal on 30 January 2015, and last updated on 6 May 2015. Additionally, copy-number calls (using GISTIC 2.0) from the cBioPortal[62] and 450k- as well as CNV-NMF-clustering results from the Broad GDAC Firehose (http://gdac.broadinstitute.org/) were acquired. Only IDH mutant samples which clearly clustered to either the 1p/19q codeleted or 1p/19q non-codeleted group (and had the respective copy-number profile; 194 samples: 124 non-codeleted, 70 codeleted) were kept for further analysis. The rationale to restrict the primary analysis on *IDH* mutated gliomas was that the *IDH* mutation itself has a profound impact on epigenetic and gene expression patterns in gliomas[2,3]. First, normalization and differential gene expression analysis of RNA sequencing counts was performed using the edgeR package[63], which assumes a negative binomial distribution of count data, filtering lowly expressed transcripts. Differentially activated signalling pathways and downstream effects between codeleted and non-codeleted IDH mutated tumours were analysed with the proprietary Ingenuity Pathway Analysis (Qiagen) using a fold change filter of |1.5| and FDR-q < 0.05[64] for the input list. Briefly, the software calculates both an overlap P value (based on Fisher's exact test) and an activation z score, which is based on the expression state of activating and inhibiting genes, for manually curated pathways and downstream biological functions. For this exploratory, hypothesis-generating study, results with both $P < 0.1$ and a z score $> |1.5|$ were kept.

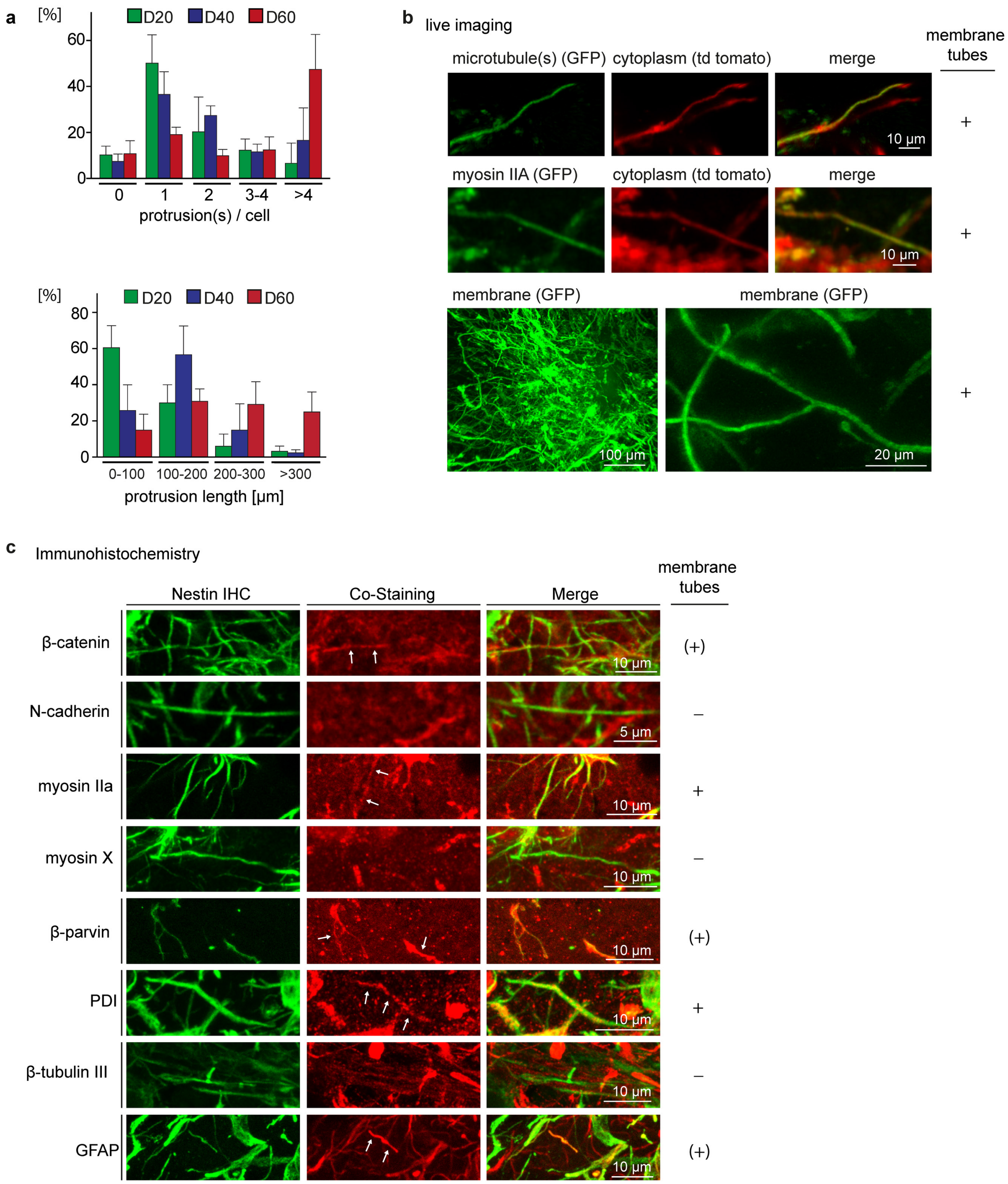
To confirm the relevance of the results for *IDH* wild-type astrocytomas, we also analysed functional transcriptomic differences between *IDH* wild-type, non-codeleted gliomas ($n = 56$) and *IDH* mutated, 1p/19q codeleted gliomas ($n = 70$) from the TCGA RNASeq data using the analysis strategy from above. As this was a secondary, exploratory analysis, we did not perform multiple-testing adjustments for the results of our primary analysis.

Statistics. The results of image analyses were transferred to the SigmaPlot Software (Systat Software, Inc.) to test the statistical significance with the appropriate tests (data were tested for normality using the Shapiro–Wilk test and for equal variance). Statistical significance was assessed by the two-sided Student's *t*-test for normally distributed data. Otherwise a Mann–Whitney test was used for non-normal distributions. For more than two groups a one way ANOVA or an ANOVA on the ranks was performed. For contingency tables, a Fisher's exact test was used. For Kaplan–Meier survival analysis, a log rank test was performed. Results were considered statistically significant if the P value was below 0.05. Quantifications were done blinded by two independent investigators. Animal group sizes were as low as possible and empirically chosen, and longitudinal measurements allowed a reduction of animal numbers by maintaining an adequate power. No statistical methods were used to predetermine sample size. If treatments were applied, animals were randomized to these procedures. Quantitative *in vivo* data are normally depicted as mean ± standard deviation. The calculated calcium imaging frequency and synchronicity values were corrected for outliers using the Nalimov test.

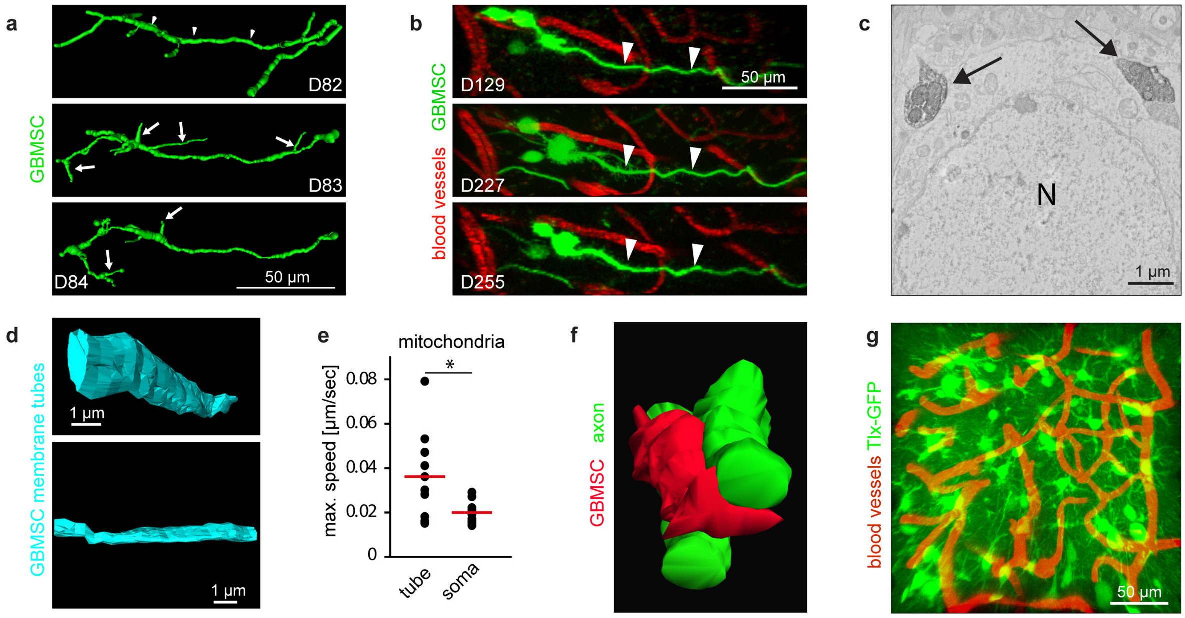
49. Winkler, F. *et al.* Kinetics of vascular normalization by VEGFR2 blockade governs brain tumor response to radiation: role of oxygenation, angiopoietin-1, and matrix metalloproteinases. *Cancer Cell* **6,** 553–563 (2004).
50. Shigetomi, E., Kracun, S. & Khakh, B. S. Monitoring astrocyte calcium microdomains with improved membrane targeted GCaMP reporters. *Neuron Glia Biol.* **6,** 183–191 (2010).
51. Thestrup, T. *et al.* Optimized ratiometric calcium sensors for functional *in vivo* imaging of neurons and T lymphocytes. *Nature Methods* **11,** 175–182 (2014).
52. Lemke, D. *et al.* Primary glioblastoma cultures: can profiling of stem cell markers predict radiotherapy sensitivity? *J. Neurochem.* **131,** 251–264 (2014).
53. Kelly, J. J. *et al.* Oligodendroglioma cell lines containing t(1;19)(q10;p10). *Neuro-oncol*. **12,** 745–755 (2010).
54. Talasila, K. M. *et al.* EGFR wild-type amplification and activation promote invasion and development of glioblastoma independent of angiogenesis. *Acta Neuropathol.* **125,** 683–698 (2013).
55. Weiler, M. *et al.* mTOR target NDRG1 confers MGMT-dependent resistance to alkylating chemotherapy. *Proc. Natl Acad. Sci. USA* **111,** 409–414 (2014).
56. Sahm, F. *et al.* Addressing diffuse glioma as a systemic brain disease with single-cell analysis. *Arch. Neurol.* **69,** 523–526 (2012).
57. Horstmann, H., Vasileva, M. & Kuner, T. Photooxidation-guided ultrastructural identification and analysis of cells in neuronal tissue labeled with green fluorescent protein. *PLoS One* **8,** e64764 (2013).
58. Horstmann, H., Körber, C., Sätzler, K., Aydin, D. & Kuner, T. Serial section scanning electron microscopy (S3EM) on silicon wafers for ultra-structural volume imaging of cells and tissues. *PLoS One* **7,** e35172 (2012).
59. Geiger, J. R. *et al.* Patch-clamp recording in brain slices with improved slicer technology. *Pflugers Arch.* **443,** 491–501 (2002).
60. Bruna, A. *et al.* High TGFβ-Smad activity confers poor prognosis in glioma patients and promotes cell proliferation depending on the methylation of the PDGF-B gene. *Cancer Cell* **11,** 147–160 (2007).
61. Opitz, C. A. *et al.* An endogenous tumour-promoting ligand of the human aryl hydrocarbon receptor. *Nature* **478,** 197–203 (2011).
62. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. **2,** 401–404 (2012).
63. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
64. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).



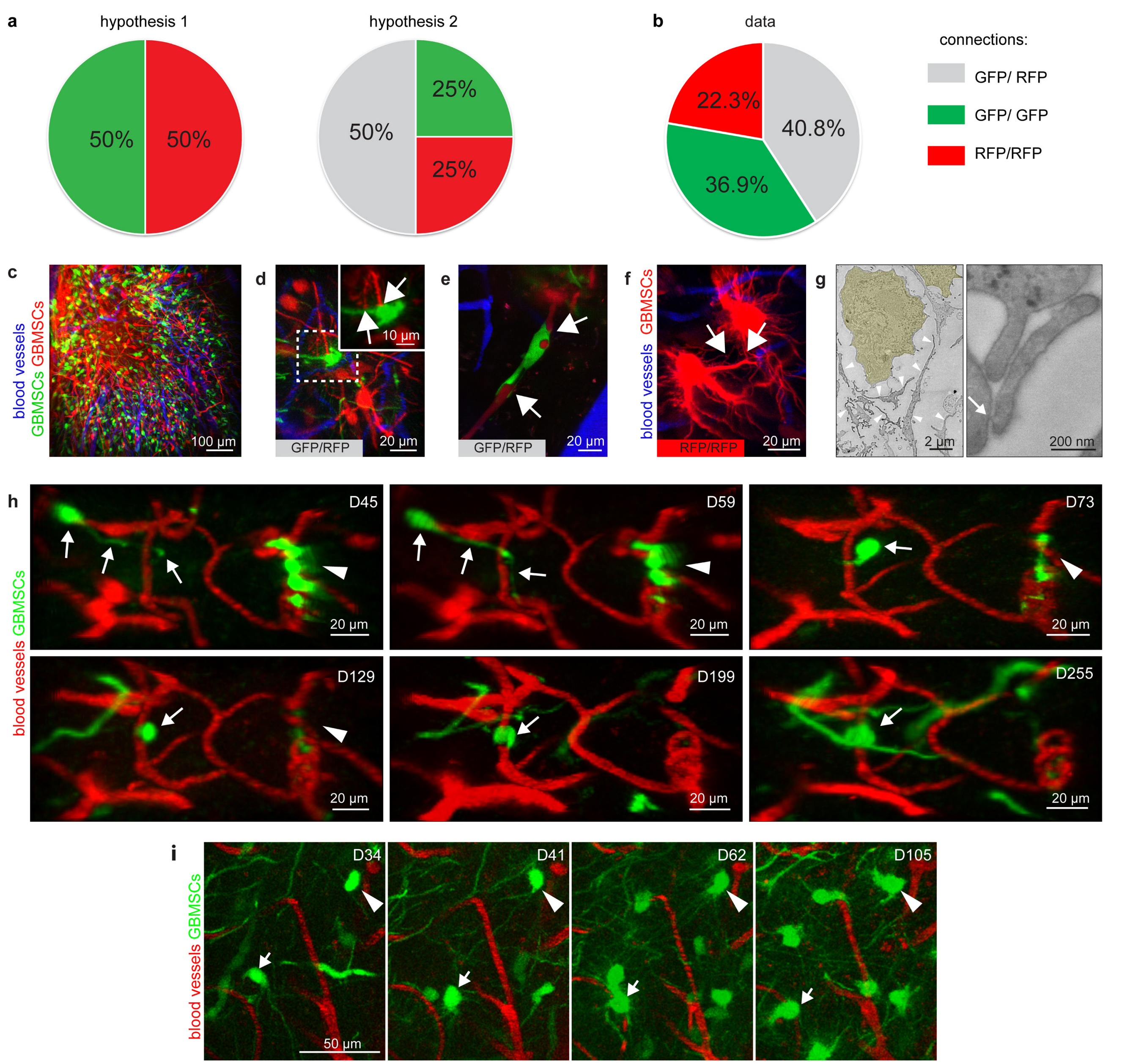
Extended Data Figure 1 | Different primary glioblastoma cell lines (GBMSCs) growing to astrocytic tumours in the mouse brain. **a**–**f**, *In vivo* microscopy (3D) of 6 different GBMSC lines (all non-codeleted for 1p/19q, and *IDH* wild-type) reveals abundant formation of ultra-long membrane protrusions in the mouse brain: T1 (**a**), T269 (**b**), T325 (**c**), S24 (**d**), WJ (**e**), and P3 (**f**) (*z*-dimensions from 200–500 µm depth). Insets show the boxed areas in the corresponding images in higher magnification, covering a proportion of the *z*-dimension. Per cell line, two time points are shown, adapted to their growth speed *in vivo* (T269, P3 fast; T1, S24 intermediate, T325 and WJ slow). **g**, 3D image of a S24 astrocytoma (injection of a 1:1 mixture of either GFP- or RFP-positive cells), revealing multiple ultra-long and very thin membrane protrusions (arrows) in the live mouse brain. Note that membrane tubes partly run in parallel. **h**, CGH-profile of the S24 GBMSC line demonstrating chromosomal alterations typical for GBM (chromosome 7 gain, 10 loss). **i**, Chromosome 7 FISH analysis of one S24 GBMSC in the main tumour area demonstrates polyploidy: 90% of $n = 100$ analysed cells in the main tumour area were clearly polyploid for chromosome 7, indicating that implanted S24 GBMSCs give rise to tumours genetically identifiable as glioblastomas. **j**, Whole mouse brain coronar sections at day 171 after S24 injection showing two main features of glioblastoma growth: diffuse brain invasion in a typical dissemination pattern (left image), and a solid, angiogenic core identified by haemorrhagic changes of the main tumour area (right bright field image). **k**, Increasing angiogenesis in this tumour is further demonstrated by dynamic *in vivo* MPLSM. **l**, Actin-rich S24 GBMSC tip, invading into the brain (single plane images; schematic drawing below). *In vivo* MPLSM: **a**–**g**, **k**, **l**.



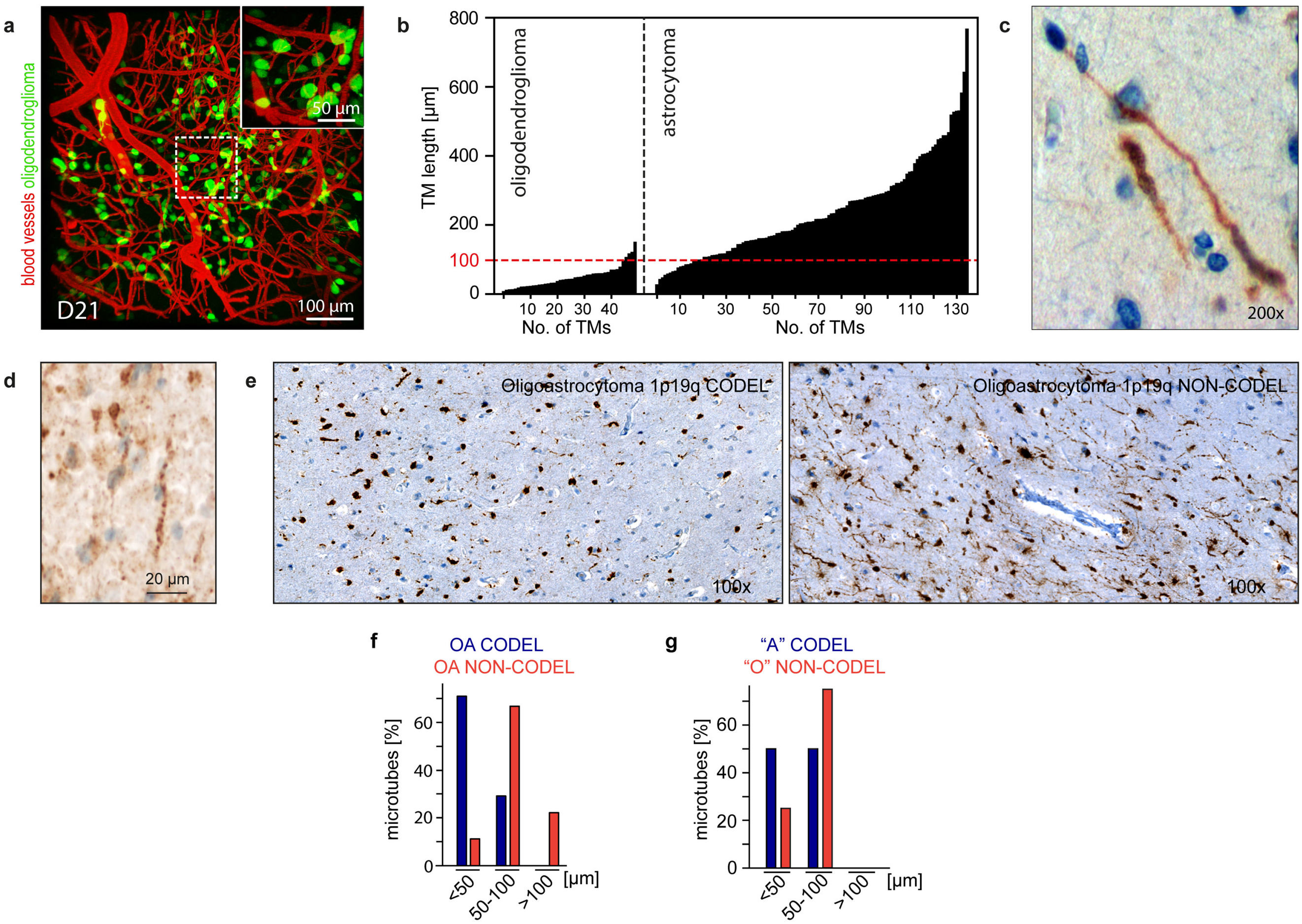
Extended Data Figure 2 | Characterization of membrane microtubes in astrocytoma mouse models. a, Number and length of protrusions during tumour progression (S24 tumours; $n = 77$–120 cells in $n = 3$ mice). **b**, MPLSM images of S24 GBMSCs genetically expressing green fluorescent protein (GFP, green) linked to different cellular/molecular components. **c**, Confocal immunohistochemistry (maximum intensity projections) of human nestin (green, allows specific detection of S24 GBMSC-related structures in the mouse brain), and different other cellular and molecular factors (red, co-stainings). The degree of expression of the factor in tumour cell-derived membrane tubes is indicated in the right lane. −, no signal in membrane tubes, (+), positive signal in some membrane tubes, +, positive signal in all membrane tubes. *In vivo* MPLSM, **a**, **b**.



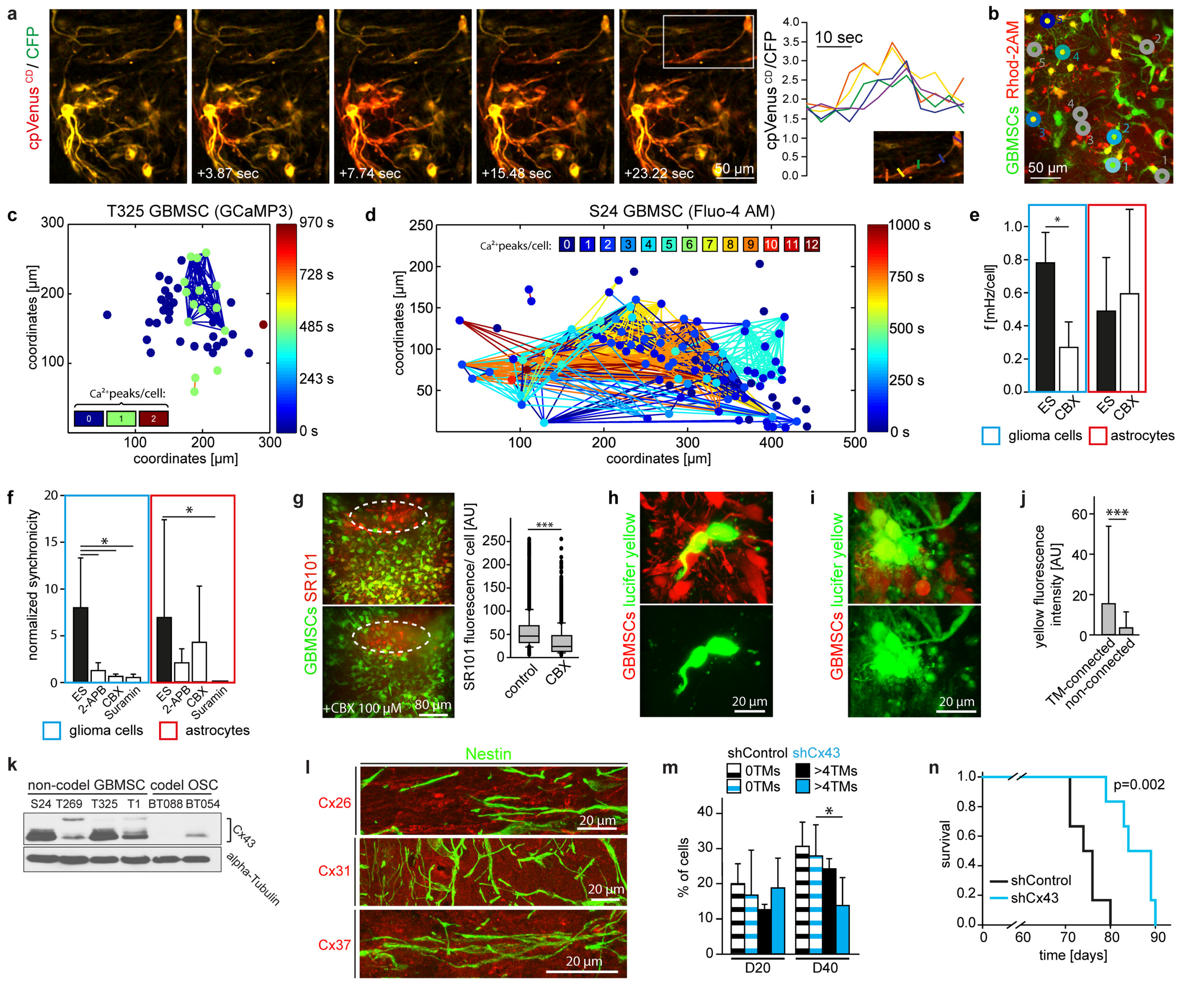
Extended Data Figure 3 | Membrane microtube dynamics and morphology. **a**, 3D reconstruction of membrane microtubes in a T325 astrocytoma over 3 days (*in vivo* MPLSM). Arrowheads, stable main tube; arrows, dynamic side tubes. **b**, Example of a very stable T325 GBMSC membrane microtube (arrowheads), followed over 126 days *in vivo*; MIP, z-dimension 48 µm. **c**, Scanning electron microscopy (SEM) image of two photoconverted membrane microtubes (arrows) and a nucleus of a non-photoconverted brain cell (N). **d**, 3D reconstruction of serial SEM images (22.29 µm (xy) × 4.62 µm (z) = 102.9 µm^3) illustrating the membrane contours. **e**, Maximum speed of mitochondria in S24 membrane tubes versus tumour cell soma ($n = 10$ per group, t-test, red lines show means). **f**, 3D reconstruction of serial SEM sections of the membrane microtube (red) and the two axons (green), which are shown in Fig. 1f. **g**, 3D image of the genetic Tlx mouse glioma model, with abundant membrane microtubes connecting single stem-like astrocytoma cells (z-dimension 83 µm). *In vivo* MPLSM: **a**, **b**, **e**, **g**. $*P < 0.05$.



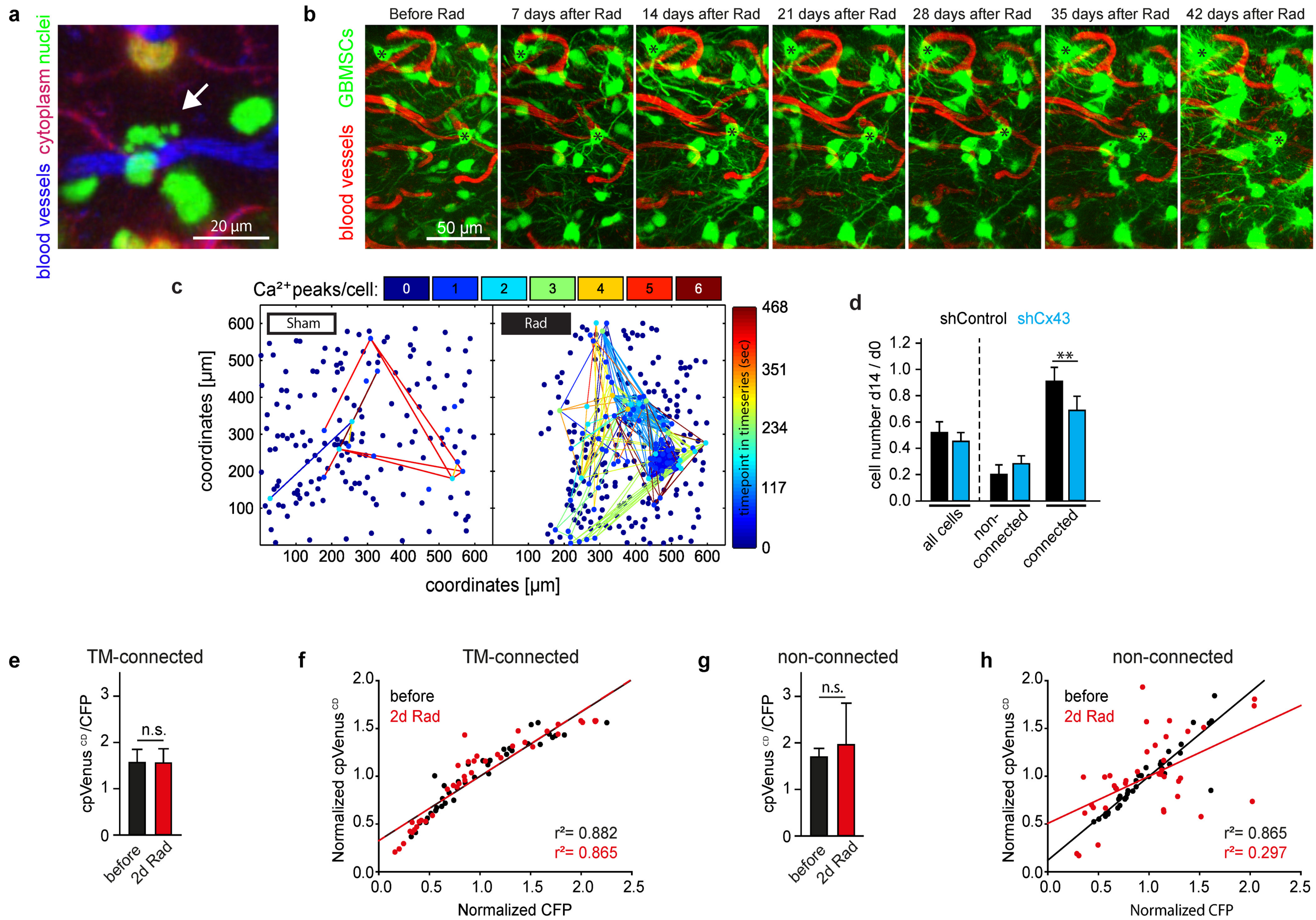
Extended Data Figure 4 | Origin of TM-connections between astrocytoma cells, and long-time tracking of TM-extending cells. **a**, Graphs illustrating two theoretically possible ways of intercellular connections by membrane tubes in a model of two tumour cell populations marked with 2 different fluorescent proteins. In hypothesis 1, tumour cells remain connected after cell division with their ancestors. In this case, only connections between cells of the same colour are expected (GFP–GFP (green) or RFP–RFP (red)). In hypothesis 2, tumour cells only connect to unrelated glioma cells. Here, 50% of connections would be between cells of different colour (GFP–RFP or RFP–GFP (grey)), and 25% of the same colour (GFP–GFP (green) and RFP–RFP (red)), respectively. **b**, Quantification of the real data set, where a 1:1 mixture of either GFP or RFP expressing S24 GBMSCs (S24GFP/S24RFP) was co-injected into the mouse brain, revealing that both potential mechanisms are in place ($n = 164$ connections in $n = 3$ mice). **c**, 3D image (70 days after injection) of a co-implantation of GFP- and RFP-expressing S24 GBMSCs. Quantification revealed that both large fluorophores (which cannot pass gap junctions) never colocalized in cell somata or in TMs ($n \geq 2{,}500$ astrocytoma cells analysed). **d**, **e**, Examples of 3D images of membrane tube connections between individual, non-related astrocytoma cells that differently express GFP or RFP (arrows in **d** and **e**). **f**, Example of a 3D image of same-colour connections between two RFP-positive cells (arrows). **g**, Scanning electron microscopy image of a S24 spheroid. Left, yellow colour marks cell bodies, arrowheads point to membrane microtubes; right, high magnification of tubes with direct membrane contact (arrow). **h**, 3D images of a perivascular T325 astrocytoma cell (arrows), which first utilizes a TM to explore the perivascular niche (D45–D73) until it moves to the explored region, and remains in a strict perivascular position until day 255. A second cell (arrowhead) is quiescent until D129 and is embedded into a vascular loop formation, which persists after disappearance of the main cell soma. **i**, MIP of a TM-containing S24 GFP astrocytoma cell which enters a perivascular position over time (arrow), and another one which remains in its non-vascular (parenchymal) position over 105 days (arrowhead). *In vivo* MPLSM, **c**–**f**, **h**, **i**; 50–650 µm deep in the brain.



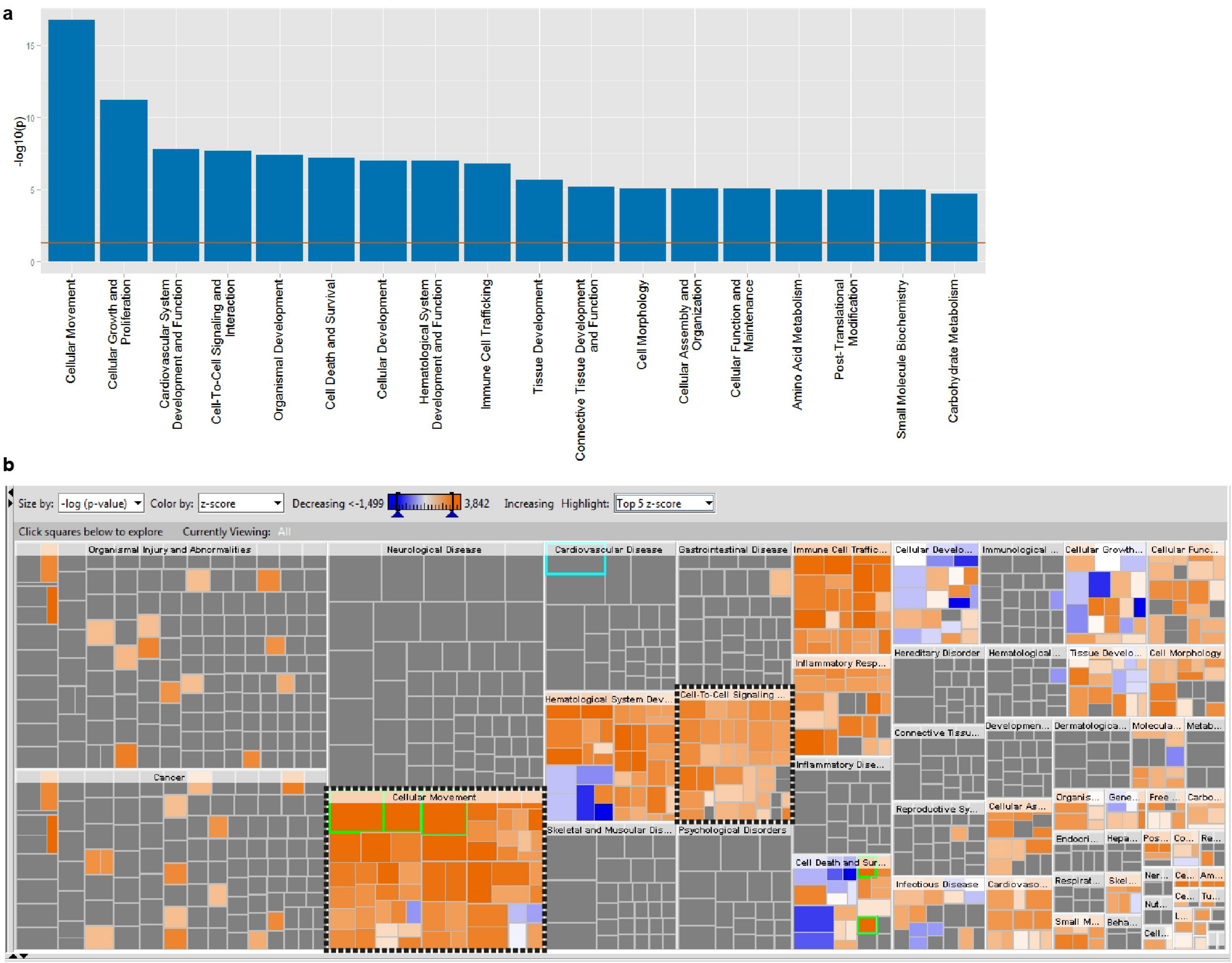
Extended Data Figure 5 | TMs in 1p/19q codeleted versus non-codeleted gliomas. **a**, 3D image (*in vivo* MPLSM) of a BT088 oligodendroglioma xenograft tumour growing in the mouse brain, inset shows the boxed area in a higher magnification. Cells are rounded, TMs are scarce. **b**, Quantification of TM lengths of BT088 oligodendroglioma cells (left), and S24 astrocytoma cells (right), at day 60 after tumour implantation. $n = 3$ animals per entity. **c**, $IDH1^{R132H}$ immunohistochemistry of the contralateral brain hemisphere (macroscopically tumour-free) of a patient deceased from a WHO III astrocytoma. **d**, Staining of resected primary glioblastomas ($n = 3$, non-codeleted, *IDH* wild-type) with a mutation-specific antibody against their $BRAF^{V600E}$ mutation reveals the existence of long tumour-cell-derived membrane microtubes in these tumours. Representative image. **e**, Exemplary $IDH1^{R132H}$ immunohistochemistry of gliomas morphologically classified as oligoastrocytoma, with (left) or without (right) 1p/19q codeletion. **f**, Maximum microtube length of oligoastrocytomas with 1p/19q codeletion (OA CODEL; $n = 31$ patients) and without (OA NON-CODEL; $n = 9$ patients). **g**, Maximum microtube length of tumours morphologically classified as astrocytomas but with 1p/19q codeletion ("A" CODEL; $n = 6$ patients), or classified as oligodendrogliomas but without 1p/19q codeletion ("O" NON-CODEL; $n = 9$ patients). *In vivo* MPLSM: **a**, **b**.



Extended Data Figure 6 | Intercellular communication via gap junctions in TM-connected astrocytoma cells, and its impact on tumour progression. a, Example of a calcium wave involving TMs of GBMSCs in a tumour region; measurement by the genetically encoded sensor Twitch-3 that allows ratiometric calcium measurements via FRET. Shown is an overlay of cpVenus^{CD} and CFP channels. Yellow colour reflects low, red colour high calcium concentrations. Right: ratios of single sections of one TM illustrating the propagation of a calcium wave along the TM. **b**, MIP (10 slices) of the region shown in Fig. 2b (red cells, astrocytes; green cells, tumour cells without Rhod-2AM signal; yellow cells, tumour cells with Rhod-2AM signal). **c**, Exemplary heat map of intercellular calcium wave (ICW) communications between T325 astrocytoma cells transfected with the genetically-encoded calcium sensor GCaMP3. **d**, Heat map of the region shown in Supplementary Video 4 (small molecule calcium indicator Fluo-4AM). **e**, Frequency of calcium peaks recorded during brain superfusion with extracellular saline (ES-control) versus 100 µM carbenoxolone (CBX) in GBMSCs (blue box) and normal brain astrocytes (red box); $n = 3$ mice per group; *t*-tests. **f**, Analysis of baseline-normalized synchronicity (see Methods for details) of calcium signals between S24 GBMSC glioma cells versus those between normal brain astrocytes. Different pharmacological blockers of main propagation mechanisms of ICWs were tested: inositol triphosphate was blocked by 2-APB, cellular ATP receptors by the nonselective purinergic 2 receptor antagonist suramin, and gap junctions were blocked by CBX (glioma cells, *t*-tests; astrocytes, Mann–Whitney tests). ES, extracellular saline used as control. **g**, 3D images (*z*-dimension 180 µm) of SR101 microinjected tumours, without (control, upper image) and with co-injected CBX (lower image; area of injection: circles) 120 min. after injection. Red cells, normal brain astrocytes. Graph, corresponding quantification of SR101-fluorescence ($n = 4{,}962$–$5{,}676$ cells in $n = 3$ mice per group; Mann–Whitney test). **h**, 3D images of a non-TM-connected S24 tumour cell (S24tdTomato), loaded with the gap-junction permeable dye Lucifer yellow via electroporation. **i**, 3D images of TM-connected S24 tumour cells (S24tdTomato) after dye transfer into one of the TM-connected cells. **j**, Quantification of Lucifer yellow fluorescence intensity in the neighbouring cells next to the electroporated cell ($n = 4$ sections from $n = 2$ mice; $n = 64$ TM-connected versus $n = 42$ non-TM connected cells quantified; *t*-test). **k**, Western blot analysis of Cx43 protein expression in 4 GBMSC and 2 oligodendroglioma stem-like (OSC) cell lines. **l**, Immunohistochemistry demonstrating the localization of different connexins in S24 GBMSCs; no clear TM-related expression, and/or localization at TM crossings could be observed. **m**, Proportion of TM-devoid (0 TMs) versus TM-rich (>4 TMs) cells in shControl versus shCx43 tumours 20 and 40 days after tumour implantation ($n = 3$ mice per group, ANOVA, Tukey's post hoc test). **n**, Kaplan–Meier survival plot of animals implanted with shCx43 vs. shControl S24 GBMSCs (log rank test). **a**–**g**, **m**, Acquired by *in vivo* MPLSM. **h**, **i**, **l**, Confocal microscopy images. For gel source data, see Supplementary Fig. 1. Scale bars show s.d. $^{*}P < 0.05$, $^{***}P < 0.001$.



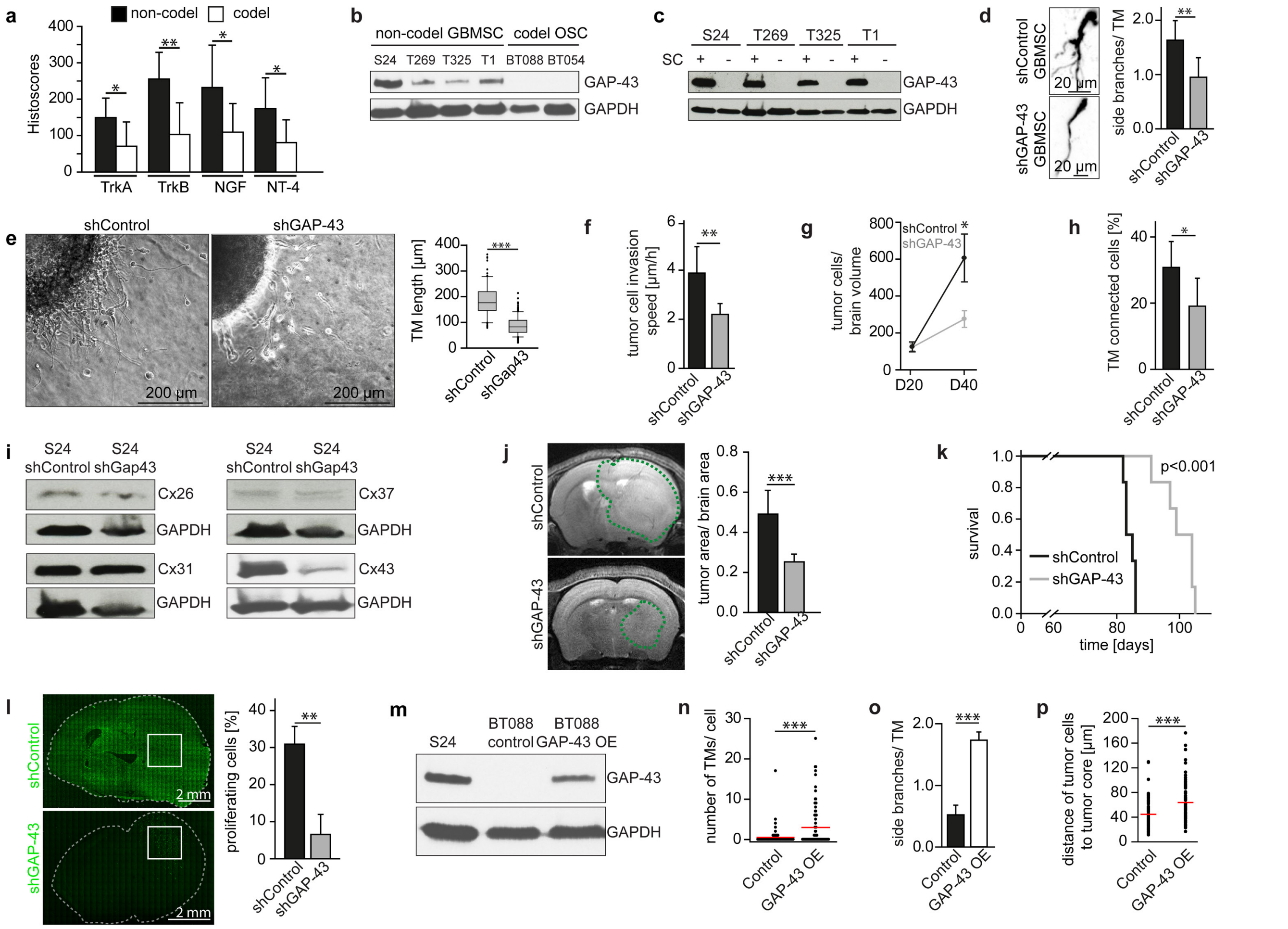
Extended Data Figure 7 | Effects of radiotherapy on cellular morphology, long-term survival, tumour cell communication, and calcium homeostasis in astrocytomas. **a**, 5 days after initiation of radiotherapy (3×7 Gy), nuclear fragmentation characteristic for apoptosis (arrow) can be detected in a proportion of cells. Green, nuclear staining by H2B–GFP transduction; red, S24 cell cytoplasm. **b**, Representative 42 day time course of a distinct tumour microregion, followed after start of radiotherapy (day 0). TM-connected cells (two examples are marked with black asterisks) show long-term survival; note that surviving cells show an increase in the number of their TMs. $n = 3$ mice per group. **c**, Exemplary heat maps of calcium transients (Rhod-2AM) of a sham treated (left) and radiated GBMSC tumour region (right). **d**, Relative changes of all cells (left) and subgroups of TM-connected versus non-connected GBMSCs of shControl versus shCx43 tumours after sham/radiotherapy ($n = 3$ mice per group, *t*-tests). **e**–**h**, Ratiometric measurements of basal calcium levels *in vivo*. **e**, Mean ratios of fluorescence intensities of the FRET partners cpVenus^{CD} and CFP, before, and after two days of radiation (2 × 7 Gy) in TM-connected cells ($n = 3$ mice per group; Mann–Whitney test). **f**, Fluorescence intensities (normalized by the mean intensities of the corresponding data sets) in TM-connected cells for the two FRET partners illustrated by a scatter blot (black dots represent analysed cells at the day before radiotherapy, red dots 2 days after initiation of radiotherapy); linear regression revealed similar correlation strengths at the two time points ($n = 3$ mice), reflecting very homogenous calcium levels in the astrocytoma cells before and after radiotherapy. **g**, Mean ratios of fluorescence intensities of the FRET partners before and after two days of radiation (2 × 7 Gy) in non-connected cells, $n = 3$ mice; Mann–Whitney test. **h**, Normalized fluorescence intensities in non-TM-connected cells for the two FRET partners. Here, linear regression revealed highly homogeneous basal calcium levels only before radiotherapy, while after radiotherapy the linear correlation was lost, illustrating heterogeneous calcium levels in the analysed cells. ($n = 3$ mice per group). All data acquired with *in vivo* MPLSM. GBMSCs, S24 cell line.



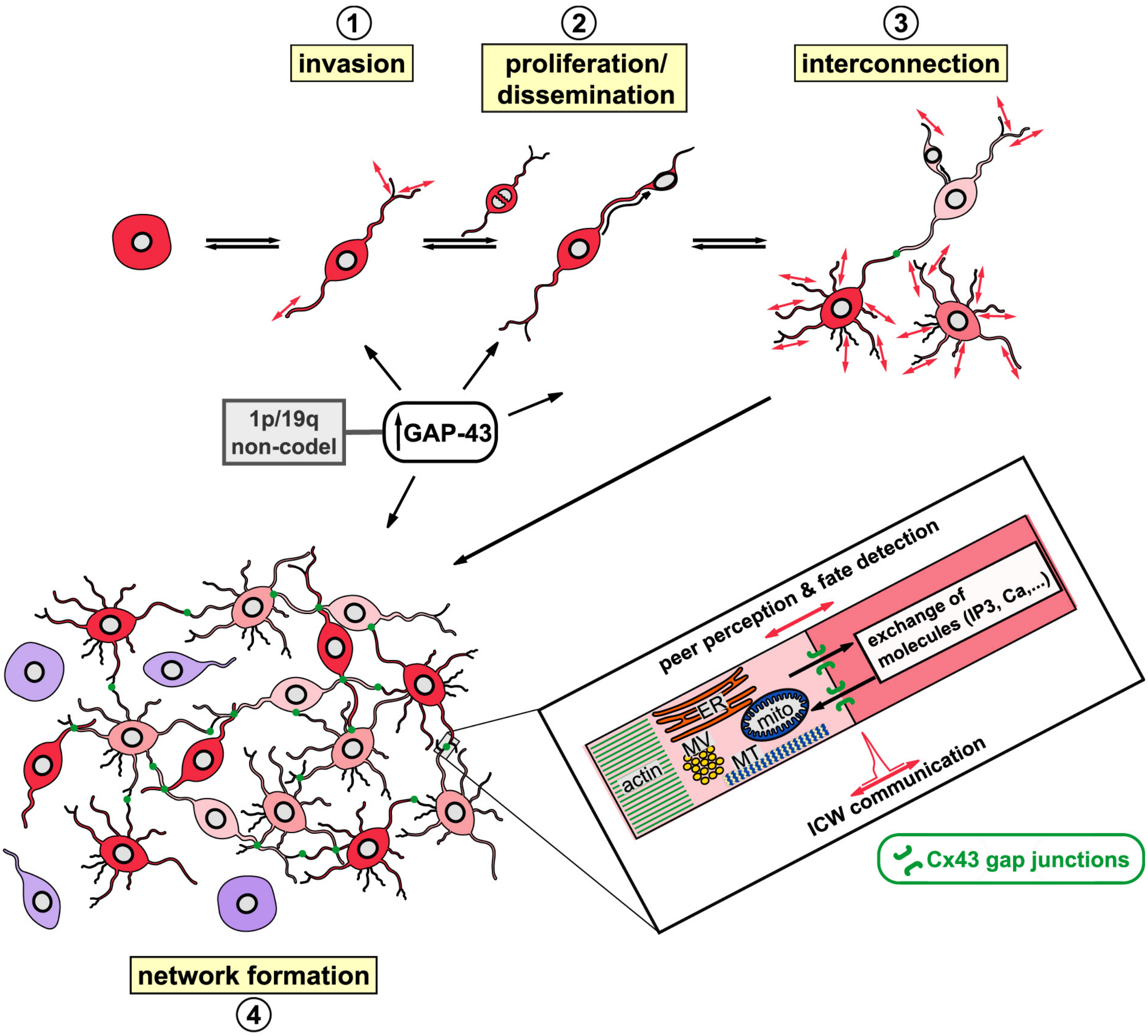
c

Ingenuity Canonical Pathways	-log(p-value)	Ratio	z-score	Molecules
Fcγ Receptor-mediated Phagocytosis in Macrophages and Monocytes	1,95E+00	1,18E-01	3,317	CDC42,SYK,INPP5D,VASP,PLD4,FYB,AKT2,ARPC1B,FCGR2A,LYN,FCGR3A/FCGR3B
IL-8 Signaling	3,88E+00	1,26E-01	3,3	JUN,IQGAP1,GNB4,MTOR,ITGB5,PLCB2,VASP,PGF,GNB1,PIK3C2B,RHOC,PLD4,ITGB2,AKT2,VCAM1,GNAI3,FOS,NRAS,GNG5,GNG4,NFKBIB,GNG12,CYBB
Acute Phase Response Signaling	2,38E+00	1,07E-01	3,207	JUN,MTOR,TF,C3,TNFRSF1A,TNFRSF1B,SERPINA3,C1R,AKT2,FOS,A2M,SOD2,NRAS,C4A/C4B,FN1,NFKBIB,FTL,C1S
Integrin Signaling	2,74E+00	1,08E-01	3,13	CDC42,ITGB5,VCL,VASP,PIK3C2B,RHOC,PARVA,PDGFB,ITGB2,ACTN1,RAP1A,AKT2,ARPC1B,NRAS,CAV1,CAPN5,ACTN2,PAK4,TNK2,ITGB4,CAPNS1
Phospholipase C Signaling	1,61E+00	8,66E-02	2,887	GNB4,PLCB2,GNB1,RHOC,PLD4,RAP1A,FCGR2A,RAPGEF3,NRAS,SYK,GNG5,GNG4,ARHGEF1,GNG12,HDAC4,HDAC9,HDAC1,ADCY8,AHNAK,LYN
Role of NFAT in Regulation of the Immune Response	4,08E+00	1,33E-01	2,887	JUN,GNB4,PLCB2,GNB1,PIK3C2B,AKT2,FCGR2A,GNAI3,HLA-DRA,FOS,NRAS,SYK,HLA-DRB1,GNG5,HLA-DMA,GNG4,HLA-DMB,NFKBIB,GNG12,CD4,LYN,FCGR3A/FCGR3B
Signaling by Rho Family GTPases	1,84E+00	9,01E-02	2,84	SEPT8,JUN,IQGAP1,CDC42,GNB4,MSN,GNB1,PIK3C2B,RHOC,ARPC1B,GNAI3,GFAP,VIM,FOS,GNG5,CYFIP1,GNG4,ARHGEF1,GNG12,PAK4,CYBB
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	2,44E+00	1,06E-01	2,828	JUN,TNFRSF1A,TNFRSF1B,APOD,PIK3C2B,JAK1,RHOC,SPI1,PPP2R2C,RAP1A,AKT2,FOS,TLR4,APOC2,PPP2R5A,NFKBIB,APOE,CYBB,CYBA
HMGB1 Signaling	1,60E+00	1,02E-01	2,714	FOS,TLR4,JUN,NRAS,CDC42,TGFB1,TNFRSF1A,TNFRSF1B,PIK3C2B,RHOC,AKT2,VCAM1
CXCR4 Signaling	2,89E+00	1,19E-01	2,673	JUN,GNB4,PLCB2,GNB1,PIK3C2B,RHOC,EGR1,AKT2,GNAI3,FOS,NRAS,GNG5,GNG4,GNG12,CD4,PAK4,ADCY8,LYN
TNFR1 Signaling	1,91E+00	1,49E-01	2,646	FOS,JUN,CDC42,TNFRSF1A,NFKBIB,CASP9,PAK4
Complement System	4,04E+00	2,50E-01	2,646	C1QC,C4A/C4B,C3,C1QB,ITGB2,C1S,C1R,C3AR1,C1QA
PKCθ Signaling in T Lymphocytes	1,08E+00	8,85E-02	2,53	FOS,JUN,NRAS,HLA-DRB1,HLA-DMA,HLA-DMB,NFKBIB,PIK3C2B,CD4,HLA-DRA
Rac Signaling	1,30E+00	9,71E-02	2,53	JUN,IQGAP1,NRAS,CDC42,CYFIP1,PIK3C2B,PAK4,CYBB,ARPC1B,CD44
MIF Regulation of Innate Immunity	1,67E+00	1,46E-01	2,449	FOS,TLR4,JUN,CD74,CD14,NFKBIB
iNOS Signaling	2,12E+00	1,63E-01	2,449	FOS,TLR4,JUN,HMGA1,CD14,NFKBIB,JAK1
IL-6 Signaling	1,02E+00	8,62E-02	2,333	FOS,JUN,A2M,NRAS,TNFRSF1A,CD14,TNFRSF1B,NFKBIB,PIK3C2B,AKT2
PI3K Signaling in B Lymphocytes	1,48E+00	9,76E-02	2,309	FOS,TLR4,JUN,NRAS,SYK,INPP5D,C3,PLCB2,NFKBIB,PLCL2,AKT2,LYN
Calcium-induced T Lymphocyte Apoptosis	1,04E+00	1,03E-01	2,236	HLA-DRB1,HLA-DMA,HLA-DMB,CD4,HDAC1,HLA-DRA
Type I Diabetes Mellitus Signaling	1,56E+00	1,04E-01	2,236	HLA-E,HLA-DRB1,HLA-DMA,TNFRSF1A,TNFRSF1B,PTPRN,HLA-DMB,NFKBIB,CASP9,JAK1,HLA-DRA
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	1,25E+00	9,24E-02	2,121	TLR4,C1QC,SYK,C3,TGFB1,IRF3,C1QB,PIK3C2B,RIPK2,C3AR1,C1QA
LPS-stimulated MAPK Signaling	1,38E+00	1,10E-01	2,121	FOS,TLR4,JUN,NRAS,CDC42,CD14,NFKBIB,PIK3C2B
JAK/Stat Signaling	1,41E+00	1,11E-01	2,121	FOS,JUN,NRAS,MTOR,CDKN1A,PIK3C2B,JAK1,AKT2
Paxillin Signaling	1,79E+00	1,12E-01	2,121	NRAS,CDC42,ITGB5,VCL,ACTN2,PIK3C2B,ITGB2,PARVA,ACTN1,PAK4,ITGB4
Cdc42 Signaling	2,40E+00	1,16E-01	2,121	JUN,HLA-DPB1,IQGAP1,CDC42,HLA-E,FNBP1L,ARPC1B,HLA-DRA,FOS,HLA-DRB1,HLA-DMA,HLA-DMB,HLA-DPA1,PAK4,TNK2
TNFR2 Signaling	1,21E+00	1,43E-01	2	FOS,JUN,TNFRSF1B,NFKBIB
Glioblastoma Multiforme Signaling	1,31E+00	8,97E-02	1,941	CDC42,MTOR,PLCB2,WNT5A,WNT7B,PIK3C2B,RHOC,PDGFB,FZD7,AKT2,NRAS,CDKN1A,PLCL2
PDGF Signaling	2,09E+00	1,30E-01	1,897	FOS,JUN,NRAS,CAV1,INPP5B,INPP5D,PIK3C2B,JAK1,INPPL1,PDGFB
fMLP Signaling in Neutrophils	2,31E+00	1,21E-01	1,897	CDC42,GNB4,PLCB2,GNB1,PIK3C2B,ARPC1B,GNAI3,NRAS,GNG5,GNG4,NFKBIB,GNG12,CYBB
TGF-β Signaling	1,74E+00	1,15E-01	1,89	FOS,JUN,TGFBR2,NRAS,CDC42,TGFB1,SMAD3,BMP2,HDAC1,BMP7
ILK Signaling	4,86E+00	1,38E-01	1,877	JUN,CDC42,TNFRSF1A,PGF,LIMS2,RHOC,PARVA,ITGB2,BMP2,VIM,FOS,FN1,PPP2R5A,ACTN2,FLNC,MYH7,MTOR,ITGB5,FLNA,PIK3C2B,PPP2R2C,ACTN1,AKT2,ITGB4,PPAP2B
Colorectal Cancer Metastasis Signaling	3,52E+00	1,13E-01	1,789	JUN,GNB4,TGFB1,TNFRSF1A,PRKAR2B,WNT5A,PGF,GNB1,RHOC,FZD7,TLR4,FOS,SMAD3,GNG4,ADCY8,WNT7B,CASP9,JAK1,PIK3C2B,AKT2,TGFBR2,NRAS,GNG5,GNG12,MMP24,MMP14
p70S6K Signaling	1,60E+00	1,02E-01	1,732	NRAS,SYK,MTOR,PPP2R5A,PLCB2,PIK3C2B,JAK1,PPP2R2C,PLCL2,AKT2,GNAI3,LYN
Ceramide Signaling	2,95E+00	1,50E-01	1,732	FOS,JUN,NRAS,PPP2R5A,TNFRSF1A,TNFRSF1B,SMPD3,S1PR3,PIK3C2B,PPP2R2C,AKT2,S1PR1
Renal Cell Carcinoma Signaling	2,97E+00	1,59E-01	1,667	FOS,JUN,NRAS,CDC42,TGFB1,PIK3C2B,PDGFB,ETS1,PAK4,RAP1A,AKT2
Neurotrophin/TRK Signaling	1,17E+00	1,04E-01	1,633	FOS,JUN,NRAS,CDC42,RPS6KA1,PIK3C2B,NTRK2
EGF Signaling	1,10E+00	1,07E-01	1,633	FOS,JUN,MTOR,PIK3C2B,JAK1,AKT2
IL-2 Signaling	1,64E+00	1,32E-01	1,633	FOS,JUN,NRAS,SYK,PIK3C2B,JAK1,AKT2
IL-1 Signaling	2,02E+00	1,21E-01	1,633	FOS,JUN,GNB4,GNG5,PRKAR2B,GNG4,NFKBIB,GNB1,GNG12,ADCY8,GNAI3
α-Adrenergic Signaling	2,24E+00	1,29E-01	1,633	NRAS,GNB4,SLC8A3,GNG5,PRKAR2B,GNG4,GNB1,GNG12,PHKG1,ADCY8,GNAI3
Cardiac Hypertrophy Signaling	1,54E+00	8,64E-02	1,5	JUN,GNB4,MTOR,TGFB1,PLCB2,PRKAR2B,GNB1,PIK3C2B,RHOC,CACNA1A,GNAI3,NRAS,TGFBR2,GNG5,GNG4,RPS6KA1,GNG12,PLCL2,ADCY8
Endothelin-1 Signaling	1,72E+00	9,52E-02	1,5	JUN,EDNRA,PLCB2,EDNRB,CASP9,PIK3C2B,GUCY1A3,PLD4,GNAI3,PRDX6,FOS,NRAS,RARRES3,LCAT,PLCL2,ADCY8
Leukocyte Extravasation Signaling	3,14E+00	1,14E-01	1,5	CDC42,MLLT4,VCL,SIPA1,CD99,VASP,MSN,PIK3C2B,RAP1GAP,ITGB2,ACTN1,RAP1A,CD44,VCAM1,RAPGEF3,GNAI3,SELPLG,ACTN2,CYBB,CYBA,MMP24,MMP14
p53 Signaling	1,09E+00	9,18E-02	-1,633	JUN,CDKN1A,PIK3C2B,GADD45B,GADD45A,HDAC9,AKT2,HDAC1,TNFRSF10B
RhoGDI Signaling	1,35E+00	8,72E-02	-2,333	CDC42,GNB4,MSN,GNB1,RHOC,ARPC1B,CD44,GNAI3,ARHGDIB,GNG5,GNG4,ARHGEF1,GNG12,PAK4,WASF2
PPAR Signaling	1,28E+00	1,00E-01	-2,828	FOS,JUN,NRAS,PPARGC1A,TNFRSF1A,TNFRSF1B,NFKBIB,NR2F1,PDGFB
PPARα/RXRα Activation	1,21E+00	8,48E-02	-3	JUN,ITGB5,TGFB1,PLCB2,PRKAR2B,FASN,NRAS,TGFBR2,PPARGC1A,SMAD3,NFKBIB,NR2F1,PLCL2,ADCY8

Extended Data Figure 8 | *In silico* analysis of 1p/19q codeleted versus non-codeleted *IDH* mutated human gliomas. Biological function analysis of 1p/19q non-codeleted ($n = 124$) versus 1p/19q codeleted ($n = 70$) human gliomas of the TCGA database was performed using Ingenuity Pathway Analysis. All tumours analysed were *IDH* mutated (GCIMP+). **a**, Bar plot of the top differentially regulated downstream biological functions. **b**, Heat map of downstream biological functions. The map is colour coded: more intense orange means more activation in 1p/19q non-codeleted tumours (compared to codeleted tumours), blue the other way round. Note the activation of "cellular movement" and "cell-to-cell signaling" in non-codeleted tumours. **c**, Results of the analysis of canonical pathways in 1p/19q non-codeleted versus codeleted gliomas. Higher positive z-score: upregulated in 1p/19q non-codeleted versus codeleted gliomas; higher negative z-score: upregulated in 1p/19q codeleted gliomas versus non-codeleted gliomas.



Extended Data Figure 9 | Proficiency for GAP-43 expression drives malignant features associated with TMs. a, TrkA, TrkB, NGF and NT-4 protein expression detected by immunohistochemistry in 1p/19 codeleted versus non-codeleted human gliomas ($n = 8$ each, *t*-tests, all *IDH* mutated). **b**, Western blot analysis of GAP-43 protein expression of different glioma cell lines. OSC, oligodendroglioma stem-like cell lines. **c**, GAP-43 western blot of 4 GBMSC lines cultured under non-adherent, stemlike (SC +) versus differentiating, serum-containing, adherent (SC −) conditions. **d**, *In vivo* 3D images of S24 shControl versus shGAP-43 GBMSCs (left) and quantification of TM side branches 20 days after implantation ($n = 60$ cells in $n = 5/6$ mice, *t*-test). **e**, Spheroid invasion assay from S24 shControl versus shGAP-43 cells in a gel matrix, and the corresponding quantification (*t*-test). **f**, *In vivo* tumour cell invasion distance within 24 h of S24 shControl versus shGAP-43 GBMSC tumours ($n = 3$ mice, Mann–Whitney test). **g**, *In vivo* proliferation dynamics in the main tumour area (volume of 0.037 mm^3; $n = 4$ mice, Mann–Whitney tests). **h**, Fraction of TM-connected cells at day 20 in these tumours ($n = 164$ cells in $n = 6$ mice, *t*-test). **i**, Western blot analysis of Cx26 (expressed in normal astrocytes), Cx31 and Cx37 (both located on chromosome 1p), and Cx43 protein expression in shGAP-43 GBMSCs versus shControls. Of note, the GAP-43 knockdown leads to a Cx43 protein reduction of 89%, while expression of the other connexins was not reduced. **j**, T2 MRI images of S24 shControl versus shGAP-43 tumours, 72 days after tumour implantation. Quantifications of $n = 6$ animals per group (*t*-test). **k**, Kaplan–Meier survival plot of S24 shControl versus shGAP-43 tumour-bearing mice (log rank test). **l**, Exemplary brain sections with nestin immunohistochemistry of S24 shControl versus shGAP-43 tumours 60 days after radiotherapy. Note that in shGAP-43 tumours, only small remnants of tumour cells can be detected by the tumour cell-specific staining. Regions with highest tumour cell densities (boxes) were quantified for proliferation index (Ki-67-positive cells/all cells; $n = 3$ animals; *t*-test). **m**, Overexpression of GAP-43 in BT088 oligodendroglioma cells results in protein levels similar to that in GBMSCs. **n**–**p**, GAP-43 overexpression in BT088 oligodendroglioma cells leads to an increase in TM numbers (**n**, $n = 80$ cells in $n = 3$ mice per group), more TM branches (**o**, $n = 40$ cells in $n = 3$ mice per group), and a higher invasion capacity (**p**, $n = 75$ cells in $n = 3$ mice per group; *t*-tests) 14 days after tumour injection. Scale bars show s.d. Red lines show means. *In vivo* MPLSM, **d**, **f**–**h**, **n**–**p**. For gel source data, see Supplementary Fig. 1. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.



Extended Data Figure 10 | Schematic illustration of the role of TMs in brain tumour progression. Anatomical and molecular mechanisms of TM-driven tumour dissemination and network function in astrocytomas. MV, microvesicles; mito, mitochondrion; ER, endoplasmic reticulum; MT, microtubules.

doi:10.1038/nature15765

Overflow metabolism in *Escherichia coli* results from efficient proteome allocation

Markus Basan[1,2*], Sheng Hui[1*], Hiroyuki Okano[1,3], Zhongge Zhang[3], Yang Shen[3], James R. Williamson[4] & Terence Hwa[1,3,5]

Overflow metabolism refers to the seemingly wasteful strategy in which cells use fermentation instead of the more efficient respiration to generate energy, despite the availability of oxygen. Known as the Warburg effect in the context of cancer growth, this phenomenon occurs ubiquitously for fast-growing cells, including bacteria, fungi and mammalian cells, but its origin has remained unclear despite decades of research. Here we study metabolic overflow in *Escherichia coli*, and show that it is a global physiological response used to cope with changing proteomic demands of energy biogenesis and biomass synthesis under different growth conditions. A simple model of proteomic resource allocation can quantitatively account for all of the observed behaviours, and accurately predict responses to new perturbations. The key hypothesis of the model, that the proteome cost of energy biogenesis by respiration exceeds that by fermentation, is quantitatively confirmed by direct measurement of protein abundances via quantitative mass spectrometry.

Under anaerobic conditions, organisms ranging from bacteria to mammalian cells excrete large quantities of fermentation products such as acetate or lactate. Notably, the excretion of these fermentation products occurs widely even in the presence of oxygen in fast-growing bacteria and fungi[1–4], as well as mammalian cells including stem cells, immune cells and cancerous cells[5–7]. This seemingly wasteful phenomenon, in which fermentation is used instead of the higher ATP-yielding respiration process for energy generation, is generally referred to as overflow metabolism (or the Warburg effect in the case of cancer[5–7]). Various rationalizations of overflow metabolism as well as regulatory schemes have been proposed over the years[2,5,8–20]. However, quantitative tests of the proposed hypotheses as well as systematic characterization of overflow metabolism are generally lacking.

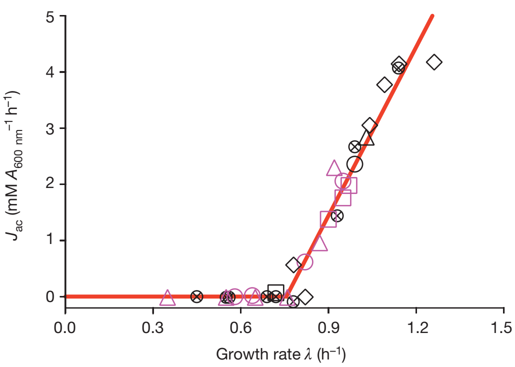
In this study, we provide a quantitative, physiological study of overflow metabolism for the bacterium *E. coli*. We report an intriguing set of linear relations between the rates of acetate excretion and steady-state growth rates for *E. coli* in different nutrient environments and different degrees of induced stresses. These relations, together with the recently established concept of proteome partition[21], led us to a simple theory of resource allocation, which can quantitatively account for all of the observed behaviours, as well as accurately predict responses to new perturbations. Key parameters of the theory regarding the proteome costs of energy biogenesis by respiration and by fermentation were determined by quantitative mass spectrometry following a coarse-graining approach. These results suggest that overflow metabolism is a programmed global response used by cells to balance the conflicting proteomic demands of energy biogenesis and biomass synthesis for rapid growth.

Threshold-linear response of acetate overflow

Previous studies have established a strong positive correlation between the rate of acetate excretion and the dilution rate for various strains of *E. coli* grown in glucose-limited continuous culture[16,22–24] (Extended Data Fig. 1a–e). Here, we measured acetate excretion and growth rates of a wild-type *E. coli* K-12 strain grown in minimal medium batch culture with a variety of glycolytic substrates as the sole carbon sources (black symbols in Fig. 1). Notably, the rate of acetate excretion per biomass, J_{ac}, exhibits a simple threshold-linear dependence on growth rate λ,

$$J_{ac} = \begin{cases} s_{ac} \cdot (\lambda - \lambda_{ac}) & \text{for } \lambda \geq \lambda_{ac} \\ 0 & \text{for } \lambda < \lambda_{ac} \end{cases} \tag{1}$$

with a linear dependence above a characteristic growth rate ($\lambda_{ac} \approx 0.76\,\text{h}^{-1}$, or 55 min per doubling), below which acetate excretion disappears. We refer to this linear relation as the acetate line (red line in Fig. 1).



Strain	Description	Carbon	Symbol
NCM3722	WT	Glucose	○
NCM3722	WT	Lactose	△
NCM3722	WT	Glycerol	□
NCM3722	WT	Other	⊗
NQ1243	Uptake titration	Glucose	○
NQ381	Uptake titration	Lactose	△
NQ636 NQ638 NQ640	*glpK* mutant	Glycerol	□
NQ3722	+ 7 AA	Various	◇

Figure 1 | Acetate excretion under carbon limitation. Acetate excretion rate (J_{ac}) is linearly correlated with the growth rate (λ) for wild-type (WT) cells grown in minimal medium with various glycolytic carbon sources (black symbols), and for cells with titratable or mutant uptake systems (purple symbols) (Extended Data Table 1). Black diamonds indicate various carbon sources supplemented with seven non-degradable amino acids (AA). The red line shows the best-fit of all the data to equation (1).

[1]Department of Physics, University of California at San Diego, La Jolla, California 92093-0374, USA. [2]Institute of Molecular Systems Biology, ETH Zürich, 8093 Zürich, Switzerland. [3]Section of Molecular Biology, Division of Biological Sciences, University of California at San Diego, La Jolla, California 92093, USA. [4]Department of Integrative Structural and Computational Biology, Department of Chemistry, The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, USA. [5]Institute for Theoretical Studies, ETH Zürich, 8092 Zürich, Switzerland.
*These authors contributed equally to this work.

For strains with titratable carbon uptake systems (Extended Data Table 1), the same linear dependence is seen for acetate excretion (Fig. 1, purple circles and triangles). These results suggest that acetate overflow is an innate response that depends on the degree of carbon influx and not specifically on the nature of carbon sources. A vivid demonstration of this effect is seen by the behaviour of cells grown on glycerol: wild-type *E. coli* cells grow in glycerol minimal medium at a rate that is below λ_{ac}, and do not excrete acetate (Fig. 1, black square), in accordance with equation (1). Three isogenic strains expressing different mutant forms of glycerol kinase[25] grew at rates faster than λ_{ac} and excreted acetate with rates dictated by their respective growth rates according to equation (1) (Fig. 1, purple squares). Instead of changing the carbon influx, reducing the metabolic demand of cells for carbon by supplementing minimal medium with non-degradable amino acids resulted in significantly enhanced growth rates and concomitantly increased acetate excretion as described by equation (1) (black diamonds in Fig. 1).

Coarse-grained model of proteome allocation

Linear growth rate dependences arose in previous physiological studies[21,26–28] from the limited capacity of ribosomes to synthesize proteins and the obligatory need for increased ribosomal proteins at faster growth[21,29]. Here, we address the problem of acetate excretion with a phenomenological resource allocation model, balancing the demand of the proteome for biomass synthesis with the demand for energy biogenesis.

We focus on acetate excretion for growth on glycolytic substrates (Fig. 1); other substrates metabolized by alternative pathways exhibit similar trends although with quantitative differences (Extended Data Fig. 1f), probably arising from the same underlying principles as those described here. In our model (detailed in Supplementary Note 1A), acetate excretion is considered as a measure of the carbon flux directed towards energy biogenesis by (oxidative) fermentation, catalysed by glycolytic enzymes and completed by the oxidative phosphorylation system (for the conversion of NADH to ATP in an aerobic environment) (Extended Data Fig. 2a). Energy biogenesis by respiration is catalysed by enzymes of the glycolysis and tricarboxylic acid (TCA) pathways, and the oxidative phosphorylation system (Extended Data Fig. 2b). Both the fermentation and respiratory pathways draw carbon flux away from biomass synthesis, via the carbon fluxes $J_{C,f}$ and $J_{C,r}$, respectively, and in turn produce the energy fluxes $J_{E,f}$ and $J_{E,r}$ (Box 1). Let the abundance of the enzymes used for fermentation and respiration be given by the fraction ϕ_f and ϕ_r, respectively, of the total protein content of the cell. All other metabolic activities, including catabolism, anabolism and ribosome synthesis (referred to as biomass synthesis), are provided by the remaining part of the proteome. Previous studies have established the growth-rate dependence of the proteome fraction for biomass synthesis[20,21,27,28], denoted here as $\phi_{BM}(\lambda)$. It is coupled to energy biogenesis via the constraint

$$\phi_f + \phi_r + \phi_{BM}(\lambda) = 1 \qquad (2)$$

The total energy flux generated must satisfy the energy demand for cell growth (denoted by $J_E(\lambda)$), that is,

$$J_{E,f} + J_{E,r} = J_E(\lambda) \qquad (3)$$

At the same time, not too much carbon should be diverted from the total influx $J_{C,in}$ order to meet the demand for biomass synthesis (flux denoted by $J_{C,BM}(\lambda)$), that is,

$$J_{C,in}(\lambda) = J_{C,f} + J_{C,f} + J_{C,BM}(\lambda) \qquad (4)$$

To a large extent, this allocation depends on the efficiencies of the energy biogenesis pathways. There are two very different efficiencies. It is well known that respiration has a much lower carbon cost—the energy flux generated per carbon is larger for respiration than fermentation[30], although this advantage of respiration is limited to the presence of oxygen (Extended Data Fig. 2). On the other hand, if respiration has a higher proteome cost, that is, if the energy flux generated per proteome fraction devoted to the respective pathway, $\varepsilon_f = J_{E,f}/\phi_f$, $\varepsilon_r = J_{E,r}/\phi_r$, is lower for respiration than fermentation, $\varepsilon_f > \varepsilon_r$, as has been suggested previously[14,19,20], then a scenario emerges that may qualitatively explain the observed disappearance of acetate flux at slow growth rates. As illustrated in Box 1, when the carbon uptake rate ($J_{C,in}$) is high and the cell has the potential to grow rapidly, it is advantageous, that is, growth rate can be maximized, to generate energy by the more proteome-efficient fermentation pathway, so that more of the proteome can be directed towards biosynthesis as required for rapid growth. Conversely, when carbon uptake is low (small $J_{C,in}$), it is advantageous to generate energy by the more carbon-efficient respiration pathway, so that more carbon flux can be directed to biosynthesis and sustain growth. This proteome allocation model predicts the carbon flux for respiration to change in the opposite way from that found for fermentation. Just as the fermentation flux can be determined from acetate excretion, the respiration flux can be deduced by measuring the rate of CO_2 evolution in a bioreactor (Supplementary Note 2). Indeed, this respiration flux exhibits a linear increase with decreasing growth rate as acetate excretion diminishes (Extended Data Fig. 3a).

Testing the model by growth perturbations

If acetate excretion is the result of the coordination of energy demand with carbon influx given constrained proteomic resources as assumed in the model, then the overexpression of useless proteins, which reduces the proteome fractions available for energy production and biomass synthesis[21], should yield higher acetate excretion rates. In fact, previous studies reported acetate excretion at slow growth rates with protein overexpression[31]. To test this hypothesis systematically, we expressed large amounts of LacZ by growing strain NQ1389 (Extended Data Table 1) on several glycolytic carbon sources. Plotting acetate excretion against growth rate for varying degrees of LacZ overexpression leads to a simple proportionality relation between growth rate and acetate excretion rate for each carbon source tested (Fig. 2a). Moreover, plotting acetate excretion against the corresponding degree of LacZ expression (fraction ϕ_Z of total cellular proteins), we find a similar linear decrease in acetate excretion rate (Extended Data Fig. 4). Finally, in a 3D plot of acetate excretion rates, LacZ abundance and growth rates (Fig. 2b), the different data points are found to lie on a single plane anchored by the acetate line (red) (see also Extended Data Fig. 4c). On this plane, acetate excretion increases linearly with LacZ overexpression at each fixed growth rate (black lines). However, for each fixed level of LacZ abundance, the plane produces a parallel shift of the standard acetate line (thin red lines). These lines are still described by equation (1), with an identical slope, but with a reduction of the threshold growth rate, λ_{ac}, linear with increasing LacZ abundance (cyan line), that is,

$$\lambda(\phi_Z) = \lambda_{ac} \cdot (1 - \phi_Z/\phi_{max}) \qquad (5)$$

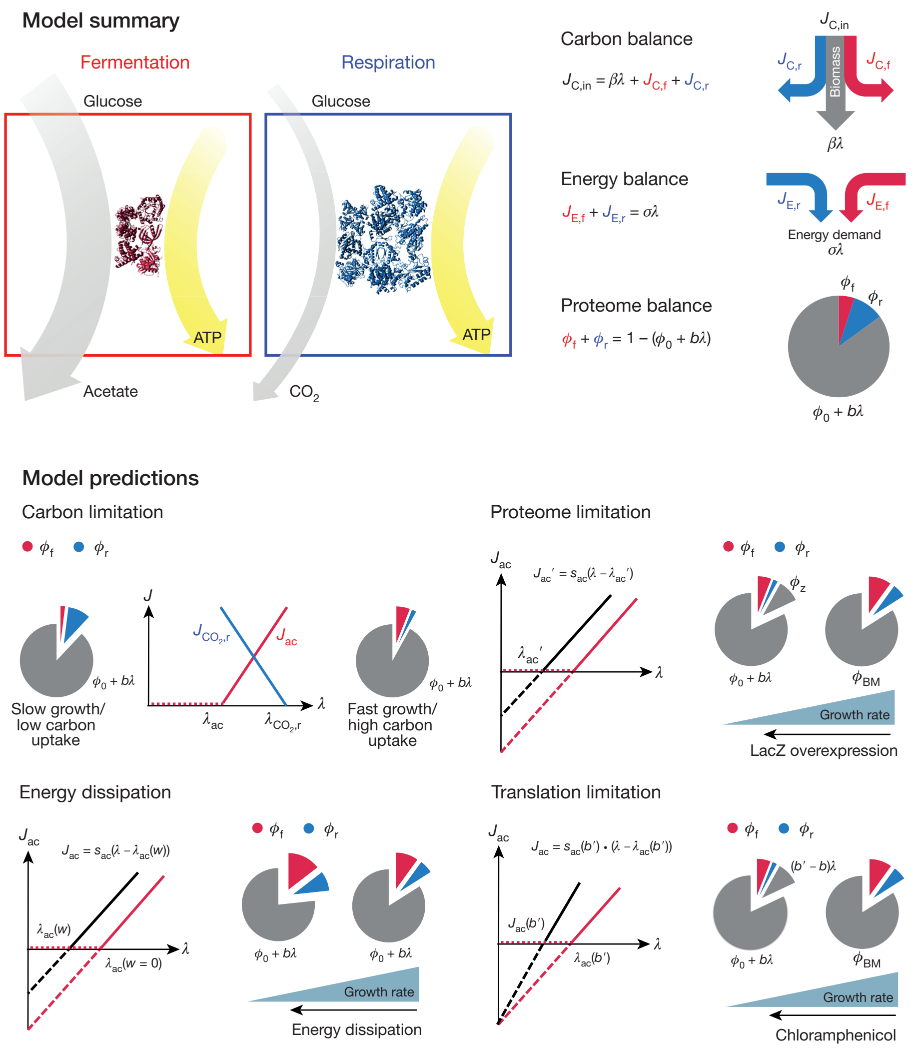
in which $\phi_{max} \approx 47\%$ is the extrapolated limit of useless protein expression at which growth rate vanishes (alternatively determined from individual lines in Extended Data Fig. 4), in agreement with previous work[21,27,28]. More quantitatively, this result is displayed in Fig. 3a, in which interpolated acetate excretion rates for constant LacZ levels are presented.

The concepts represented by equations (2)–(4) are transformed into a quantitative model (as illustrated in Box 1 and detailed in Supplementary Note 1A) by implementing a simple set of relations. First, the proteome fraction ϕ_{BM} responsible for biomass synthesis under carbon-limitation follows a linear growth-rate dependence, that is, $\phi_{BM}(\lambda) = \phi_0 + b\lambda$, as established by previous studies[21,27,28].

BOX 1

Resource allocation model for energy biogenesis

Top left, efficiencies of energy production. Fermentation and respiration pathways for energy biogenesis are shown in the red and blue boxes, respectively. The model assumes that for the same energy flux generated (width of yellow arrows), fermentation needs to draw more carbon flux than respiration (compare the width of light grey arrows), but requires smaller amount of proteins (compare the number of red and blue proteins). Top, model summary. The model consists of three resource-balance equations. (1) Carbon flux ($J_{C,in}$) is used for energy production via fermentation or respiration ($J_{C,f}$, $J_{C,r}$), and to provide precursors for biomass production ($\beta\lambda$). (2) Fermentation and respiration pathways supply ATP flux ($J_{E,f}$, $J_{E,r}$) that satisfies the energy demand of the cell ($\sigma\lambda$). (3) The proteome fraction required for biomass synthesis ($\phi_0 + b\lambda$) depends linearly on the growth rate, thereby constraining the proteome fraction available for energy biosynthesis (ϕ_f, ϕ_r). Bottom, model predictions. Under carbon limitation, the model predicts threshold-linear dependences of fermentation and respiration with changing growth rate. Respiration (blue line) gradually replaces fermentation (red line) as the growth rate decreases. Proteome limitation by expression of useless proteins results in a horizontal shift of the acetate line. Translational limitation by antibiotics results in an increased slope of the acetate line with a fixed *y*-intercept. Energy dissipation also leads to a parallel shift of the acetate line. But unlike proteome limitation, which 'compresses' both the fermentation and respiration sectors, these two sectors both increase with decreasing growth rate under energy dissipation (for fixed carbon uptake). The behaviours summarized in these plots are derived quantitatively in Supplementary Note 1 and validated in Figs 1–3.



Second, empirical evidence[28] indicates linear relations between metabolic fluxes and the abundances of the corresponding proteome sectors, which we capture by the equations $J_{C,f} = \kappa_f\phi_f$, $J_{C,r} = \kappa_r\phi_r$, $J_{E,f} = \varepsilon_r\phi_r$ and $J_{E,r} = \varepsilon_r\phi_r$. Finally, we introduce proportionalities of biomass and energy demand to the growth rate ($J_{C,BM}(\lambda) = \beta\lambda$, $J_E(\lambda) = \sigma\lambda$), relations that are demonstrated experimentally (see Supplementary Notes 1D2 and 1D4). (Maintenance energy is negligible over the growth-rate range studied[32].) The detailed meaning of each parameter introduced here is given in Extended Data Table 2. Most important among them are ϕ_f and ϕ_r, the proteome efficiencies of energy biogenesis by the fermentation and respiration pathways, respectively.

Equations (2)–(4), together with the linear relation between the proteome fractions and fluxes, describe all key features of the experimental data as detailed in Supplementary Notes 1B and 1C and illustrated in Box 1 (bottom): the model naturally gives rise to the observed threshold-linear form of acetate excretion equation (1),

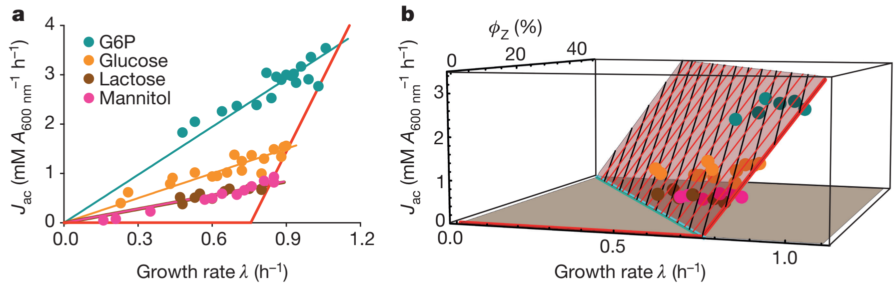


Figure 2 | Effect of protein overexpression on acetate excretion. **a**, Measured acetate excretion rate is plotted against growth rate for increasing degrees of the (useless) expression of LacZ in strain NQ1389, for several carbon sources indicated by circles of different colours. Thick red line is the acetate line of wild-type cells shown already in Fig. 1, and the thin lines are model predictions (equation (S26) in Supplementary Information), one for each carbon source of respective colour. G6P, glucose 6-phosphate. **b**, 3D plot of the data in **a**. The data lie largely on a plane spanned by the acetate line (thick red line) defined in Fig. 1, and the cyan line, $\lambda_{ac}(\phi_Z)$, which defines a linear shift in the threshold growth rate, λ_{ac}, for different degrees of LacZ expression, as predicted by the model (equation (5)).

with formulae for the threshold λ_{ac} and the slope s_{ac} given by equations (S15) and (S16) in Supplementary Information; the linear decrease in energy-related CO_2 production upon increasing growth rate (Extended Data Fig. 3a) is captured by equation (S17), with the corresponding threshold $\lambda_{CO_2,r}$ and slope $s_{CO_2,r}$ given by equations (S18) and (S19). Furthermore, the parallel shifts of the acetate line for a constant level of protein overexpression (Fig. 3a, and thin red lines in Fig. 2b) are captured by equation (S30) in Supplementary Information, while the direct proportionalities between acetate excretion rates and growth rate upon varying the degree of overexpression (thin solid lines in Fig. 2a) are captured by equation (S26). Notably, the data imposes a set of quantitative constraints on the model parameters, in particular,

$$\frac{\varepsilon_f}{\varepsilon_r} \geq \frac{\lambda_{CO_2,r}}{\lambda_{ac}} \approx 1.5, \qquad (6)$$

predicting that fermentation is at least 50% more efficient for energy biogenesis than respiration in terms of proteome cost (see equations (S20)–(S22) in Supplementary Information for a derivation).

To test quantitatively the proteome allocation model, we performed additional sets of experiments designed to perturb individual model

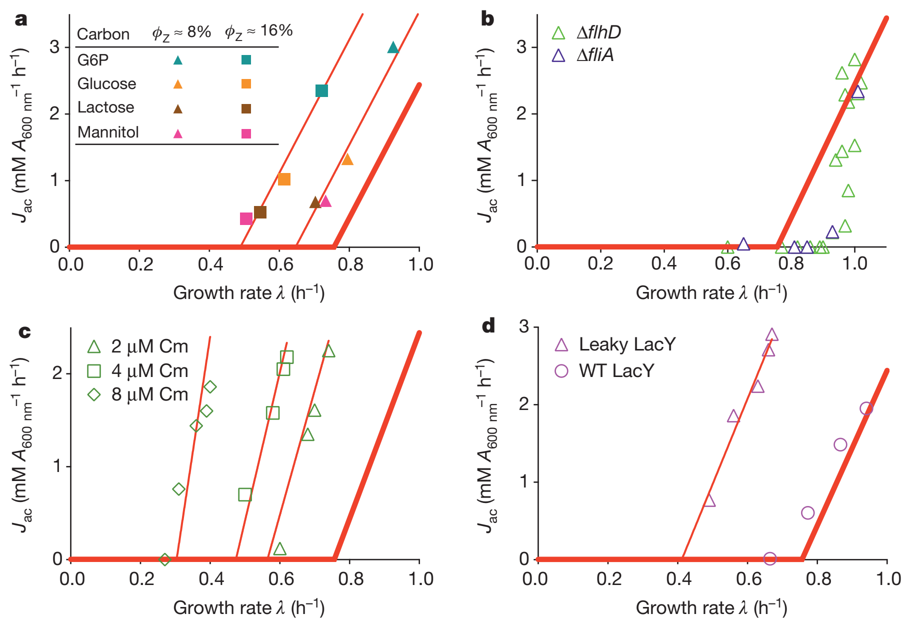


Figure 3 | Effect of genetic and environmental perturbations on the acetate line: model and experiments. In all subplots, the thick red line represents the acetate line of wild-type cells (Fig. 1). **a**, For each fixed level of LacZ expression, growth of the overexpression strain NQ1389 on different carbon sources leads to parallel shifts of the acetate line, as demonstrated by the thin red lines, whose slopes are fixed to that of the wild-type acetate line and were fitted by only adjusting the threshold growth rates λ_{ac}. **b**, Acetate excretion rate with glucose uptake titration (Pu-ptsG, Extended Data Table 1) for a Δ*flhD* strain (NQ1388) and a Δ*fliA* strain (NQ1539), both incapable of expressing motility proteins, regarded as 'useless' in well-stirred culture. Because the motility proteins are only expressed significantly as growth rate decreases under carbon limitation (Extended Data Fig. 4f), acetate excretion deviates from the acetate line as growth rate decreases. The shift of the threshold growth rate λ_{ac} in the two strains is quantitatively consistent with the model prediction (equation (5)), based on quantification of motility proteins (Extended Data Fig. 4f and Extended Data Table 3). **c**, For each fixed (sub-lethal) dose of chloramphenicol (Cm) in the medium, acetate excretion rates were determined for different degrees of lactose uptake by the titratable LacY strain (NQ381, Extended Data Table 1). The thin red lines are model fits with the slope as the only fitting parameter. **d**, For the energy dissipating mutant (NQ1313) expressing the proton-leaking LacY^{A177V}, acetate excretion rates (triangles) obtained from titrating the glucose uptake system (Extended Data Table 1) show a parallel shift of the acetate line (thin red line), obtained by fitting the data by adjusting only the threshold growth rate in accordance with model prediction (equation (S32) in Supplementary Information). For comparison, acetate excretion in cells expressing wild-type LacY from the same plasmid system (NQ1312, circles) adheres much closer to the acetate line of wild-type cells.

parameters. To see whether a decrease in acetate excretion is possible, we examined two independent mutants (Δ*flhD* and Δ*fliA*) in which motility proteins, 'useless' in well-shaken batch culture, are not expressed. These mutants exhibit reductions in acetate excretion (open symbols in Fig. 3b), in accordance with the prediction of the resource allocation model as more proteome becomes available for energy biogenesis. (The nonlinear dependence of acetate excretion arises in this case due to the growth-rate dependence of motility protein expression in wild-type cells[28], as shown in Extended Data Fig. 4f.) Translational limitation by sub-lethal doses of the antibiotic chloramphenicol inhibits peptide elongation and makes the cell respond by allocating a larger proteome fraction to ribosomes[21]. In our model, this affects the parameter b in $\phi_{BM}(\lambda)$, and therefore predicts an increased slope of the acetate line with an identical y-intercept (equation (S14) in Supplementary Information, solid lines in Fig. 3c), which is in good agreement with the data (open symbols in Fig. 3c). We also investigated the effect of energy dissipation on acetate excretion by expressing a mutant lactose transporter $\mathrm{LacY}^{\mathrm{A177V}}$ ('leaky LacY') known to leak protons across the inner membrane[33] (Extended Data Table 1). An energy leakage flux can be added to the right-hand side of equation (3), and the model predicts a parallel shift of the acetate line to higher excretion rates (equation (S32) in Supplementary Information). This prediction was tested by titrating glucose uptake in a strain (NQ1313) expressing the leaky LacY mutant. As anticipated, a parallel shift to higher acetate excretion rates was obtained (purple triangles and line in Fig. 3d). Similar increases in acetate excretion were obtained with the addition of 2,4-dinitrophenol (DNP), which uncouples oxidative phosphorylation by carrying protons across the cell membrane (Extended Data Fig. 5). A summary of quantitative comparisons between predictions of the proteome allocation model with experimental findings is presented in Extended Data Table 3, showing that the model quantitatively captures the changes of acetate excretion patterns in response to the applied perturbations.

Proteome cost of fermentation and respiration

The theoretical predictions tested so far do not require the knowledge of the values of proteome cost parameters (for example, ε_f, ε_r). However, these parameters are of central importance for theories based on proteome allocation. We have thus developed a coarse-graining approach to characterize the proteome cost for fermentation and respiration directly. First, the absolute protein abundance of individual proteins was obtained using quantitative mass spectrometry[28] together with absolute abundance calibration by ribosome profiling[34]. Next, for each enzyme involved in glycolysis, TCA and oxidative phosphorylation, its abundance was partitioned among the three pathways, fermentation, respiration and biomass synthesis, in proportion to the three fluxes through the enzyme. Finally, the fractional enzyme amounts partitioned into a pathway were summed up to obtain the total enzyme abundance devoted to the pathway (Fig. 4a; see Supplementary Note 3 for details). In Fig. 4b, the energy production fluxes of fermentation and respiration (Extended Data Fig. 3b) are plotted against their respective proteome fractions determined in this manner. The linearity of the results validates the linear dependences between $J_{E,f}$, $J_{E,r}$ and ϕ_f, ϕ_r assumed in the model, while slopes of these lines directly yield the proteome efficiency of energy biogenesis for glycolytic carbon sources: $\varepsilon_f \approx 750$ mM ATP per $A_{600\,nm}$ per hour, and $\varepsilon_f \approx 390$ mM ATP per $A_{600\,nm}$ per hour. Indeed, the proteome cost of fermentation ($1/\varepsilon_f$) is approximately twofold lower than that of respiration ($1/\varepsilon_r$), quantitatively validating the key assumption of this work. Together with similar procedures used to determine the other model parameters (as described in detail in Supplementary Note 1D), we obtained a self-consistent set of parameters (Extended Data Table 2) that successfully recapitulates all our experimental data (Extended Data Fig. 3c, d).

The proteome allocation model is able to predict not only acetate excretion patterns but also the expression of dozens of genes in the glycolysis and TCA pathways under different perturbation, as detailed in Supplementary Note 1C and Extended Data Figs 6 and 7. While both proteome limitation and energy dissipation lead to parallel shifts of the acetate line (Fig. 3a and c, respectively), this response arises from an opposite response of the energy sectors ϕ_f, ϕ_r as predicted by the model and verified by mass spectrometry. Under LacZ overexpression, cells decreased the expression of enzymes for both fermentation and respiration (orange lines in Extended Data Figs 6 and 7, as predicted in equations (S26)–(S27) in Supplementary Information), while under energy dissipation, cells increased the expression of these enzymes (blue lines in Extended Data Figs 6 and 7, predicted by equations (S36)–(S37)).

Discussion

The notion that fermentation may be more proteome efficient than respiration was proposed previously by Molenaar *et al.*[14], extended to the use of the Entner–Doudoroff pathway by Flamholz *et al.*[35] and to the genome-scale by O'Brien *et al.*[19]. Our study directly verifies this hypothesis (Fig. 4b), and establishes the pivotal role proteome efficiency has in determining the degree of overflow metabolism in *E. coli* (Fig. 3). Our findings in response to useless protein expression and energy dissipation are difficult to reconcile, even qualitatively, with alternative hypotheses such as the limitation of respiratory capacity[8], the need for recycling of cofactors[2], or constraints of the cytoplasmic membrane[17]. Models with cell volume constraints[12] are mathematically similar to protein cost models; however, cell volume varies widely between growth conditions with similar densities[36], suggesting that it is not a constraint.

Mechanistically, the re-uptake of acetate by acetyl-CoA synthase (ACS), upregulated by the cAMP receptor protein (CRP)–cAMP

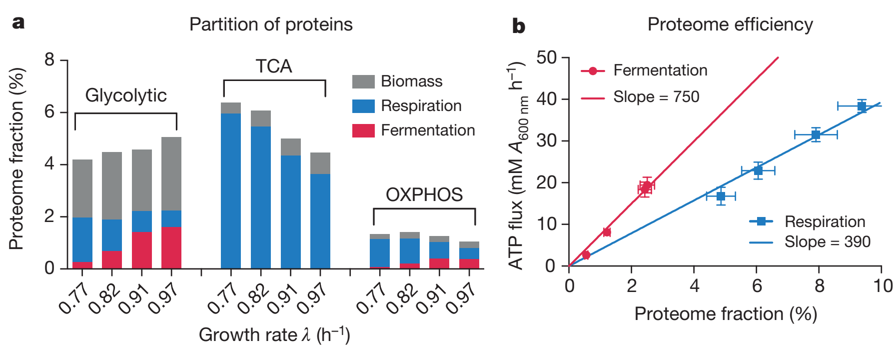


Figure 4 | Partition of proteome fractions into flux components. a, The total abundance of all proteins devoted to glycolysis, TCA and oxidative phosphorylation (OXPHOS) (defined in Supplementary Note 3), is represented as a fraction of total protein for various degrees of lactose uptake in strain NQ381. Different colours indicate the proportions of these proteins dedicated to biomass production, fermentation and respiration, estimated using the corresponding fraction of fluxes; see Supplementary Note 3. **b**, Energy fluxes through the fermentation and respiration pathways are plotted against their respective proteome fractions. The lines are linear regressions of the data, with the slopes being the proteome efficiencies (ϕ_f and ϕ_r). The steeper slope (for fermentation) indicates higher ATP production per protein devoted to the pathway (lower protein cost), validating the central assumption of the proteome allocation model.

complex under carbon limitation[27], is important for the decrease of acetate excretion under carbon limitation[15,16]. However, the linear growth-rate dependence of carbon overflow demands the tight global coordination of energy biogenesis pathways with biosynthesis, which cannot be accounted for by ACS activity alone, and requires the coordinated regulation of glycolytic and TCA enzymes. In particular, acetate excretion sharply increased under energy dissipation while the abundance of ACS proteins increased slightly as well (Extended Data Fig. 6). Moreover, the parallel increase of glycolytic and TCA enzymes, in combination with the increase in acetate excretion under energy dissipation, cannot be rationalized from the known actions of CRP regulation (Extended Data Figs 6 and 7), suggesting the role of additional regulator(s).

We have established how diverse patterns of acetate excretion can be understood as a part of a global physiological response used by *E. coli* to cope with changing proteomic demands of energy biogenesis and biomass synthesis under different growth conditions[37]. Our findings can be used to guide approaches to minimizing overflow metabolism in synthetic biology applications[24,38–41], in ways congruent with the fitness of the organism, for example, by reducing the expression of useless proteins (Fig. 3b). More broadly, a similar physiological rationale may underlie overflow metabolism in rapidly growing eukaryotes including tumour cells[7,42], in which the synthesis of mitochondria for TCA reactions is an additional cost. Indeed, scatter plots of ethanol production and sugar uptake in various strains of *Saccharomyces cerevisiae* and other yeast species point to the existence of a universal response similar to that shown for *E. coli*[18] (Fig. 1). The quantitative physiological approach developed in this work can be used as a model for characterizing metabolic efficiency and its biological implications in these systems and others.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 January; accepted 1 October 2015.

1. Neidhardt, F. C., Ingraham, J. L. & Schaechter, M. *Physiology of the Bacterial Cell: A Molecular Approach* Ch. 5 (Sinauer Associates Inc, 1990).
2. Wolfe, A. J. The acetate switch. *Microbiol. Mol. Biol. Rev.* **69,** 12–50 (2005).
3. De Deken, R. H. The Crabtree effect: a regulatory system in yeast. *J. Gen. Microbiol.* **44,** 149–156 (1966).
4. De Mey, M., De Maeseneire, S., Soetaert, W. & Vandamme, E. Minimizing acetate formation in *E. coli* fermentations. *J. Ind. Microbiol. Biotechnol.* **34,** 689–700 (2007).
5. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324,** 1029–1033 (2009).
6. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144,** 646–674 (2011).
7. Weinberg, R. A. *The Biology of Cancer* 2nd edn, Ch. 2.6 (Garland Science, 2013).
8. Majewski, R. A. & Domach, M. M. Simple constrained-optimization view of acetate overflow in *E. coli*. *Biotechnol. Bioeng.* **35,** 732–738 (1990).
9. Varma, A. & Palsson, B. O. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* **60,** 3724–3731 (1994).
10. Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and competition in the evolution of ATP-producing pathways. *Science* **292,** 504–507 (2001).
11. Pfeiffer, T. & Bonhoeffer, S. Evolutionary consequences of tradeoffs between yield and rate of ATP production. *Z. Phys. Chem.* **216,** 51–63 (2002).
12. Vazquez, A. *et al.* Impact of the solvent capacity constraint on *E. coli* metabolism. *BMC Syst. Biol.* **2,** 7 (2008).
13. Vemuri, G. N., Altman, E., Sangurdekar, D. P., Khodursky, A. B. & Eiteman, M. A. Overflow metabolism in *Escherichia coli* during steady-state growth: transcriptional regulation and effect of the redox ratio. *Appl. Environ. Microbiol.* **72,** 3653–3661 (2006).
14. Molenaar, D., van Berlo, R., de Ridder, D. & Teusink, B. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol. Syst. Biol.* **5,** 323 (2009).
15. Valgepea, K. *et al.* Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase. *BMC Syst. Biol.* **4,** 166 (2010).
16. Renilla, S. *et al.* Acetate scavenging activity in *Escherichia coli*: interplay of acetyl-CoA synthetase and the PEP-glyoxylate cycle in chemostat cultures. *Appl. Microbiol. Biotechnol.* **93,** 2109–2124 (2012).
17. Zhuang, K., Vemuri, G. N. & Mahadevan, R. Economics of membrane occupancy and respiro-fermentation. *Mol. Syst. Biol.* **7,** 500 (2011).
18. Huberts, D. H., Niebel, B. & Heinemann, M. A flux-sensing mechanism could regulate the switch between respiration and fermentation. *FEMS Yeast Res.* **12,** 118–128 (2012).
19. O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R. & Palsson, B. O. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* **9,** 693 (2013).
20. Peebo, K. *et al.* Proteome reallocation in *Escherichia coli* with increasing specific growth rate. *Mol. Biosyst.* **11,** 1184–1193 (2015).
21. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. *Science* **330,** 1099–1102 (2010).
22. Meyer, H.-P., Leist, C. & Fiechter, A. Acetate formation in continuous culture of *Escherichia coli* K12 D1 on defined and complex media. *J. Biotechnol.* **1,** 355–358 (1984).
23. Nanchen, A., Schicker, A. & Sauer, U. Nonlinear dependency of intracellular fluxes on growth rate in miniaturized continuous cultures of *Escherichia coli*. *Appl. Environ. Microbiol.* **72,** 1164–1172 (2006).
24. el-Mansi, E. M. & Holms, W. H. Control of carbon flux to acetate excretion during growth of *Escherichia coli* in batch and continuous cultures. *J. Gen. Microbiol.* **135,** 2875–2883 (1989).
25. Applebee, M. K., Joyce, A. R., Conrad, T. M., Pettigrew, D. W. & Palsson, B. O. Functional and metabolic effects of adaptive glycerol kinase (GLPK) mutants in *Escherichia coli*. *J. Biol. Chem.* **286,** 23150–23159 (2011).
26. Scott, M. & Hwa, T. Bacterial growth laws and their applications. *Curr. Opin. Biotechnol.* **22,** 559–565 (2011).
27. You, C. *et al.* Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* **500,** 301–306 (2013).
28. Hui, S. *et al.* Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol.* **11,** 784 (2015).
29. Maaloe, O. in *Biological Regulation and Development* Vol. 1 (ed. Goldberger, R. F.) 487–542 (Plenum, 1979).
30. Nelson, D. L., Lehninger, A. L. & Cox, M. M. *Lehninger Principles of Biochemistry* Chs 14, 16 (Macmillan, 2008).
31. Sandén, A. M. *et al.* Limiting factors in *Escherichia coli* fed-batch production of recombinant proteins. *Biotechnol. Bioeng.* **81,** 158–166 (2003).
32. Neijssel, O. M., Teixeira de Mattos, M. J. & Tempest, D. W. in Escherichia coli *and* Salmonella*: Cellular and Molecular Biology* (eds Neidhardt, F. C. *et al.*) 1683–1693 (ASM Press, 1996).
33. Brooker, R. J. An analysis of lactose permease "sugar specificity" mutations which also affect the coupling between proton and lactose transport. I. Val177 and Val177/Asn319 permeases facilitate proton uniport and sugar uniport. *J. Biol. Chem.* **266,** 4131–4138 (1991).
34. Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157,** 624–635 (2014).
35. Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W. & Milo, R. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl Acad. Sci. USA* **110,** 10039–10044 (2013).
36. Woldringh, C. L., Binnerts, J. S. & Mans, A. Variation in *Escherichia coli* buoyant density measured in Percoll gradients. *J. Bacteriol.* **148,** 58–63 (1981).
37. Bollenbach, T., Quan, S., Chait, R. & Kishony, R. Nonoptimal microbial response to antibiotics underlies suppressive drug interactions. *Cell* **139,** 707–718 (2009).
38. El-Mansi, M. Flux to acetate and lactate excretions in industrial fermentations: physiological and biochemical implications. *J. Ind. Microbiol. Biotechnol.* **31,** 295–300 (2004).
39. Farmer, W. R. & Liao, J. C. Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nature Biotechnol.* **18,** 533–537 (2000).
40. Aristidou, A. A., San, K. Y. & Bennett, G. N. Metabolic engineering of *Escherichia coli* to enhance recombinant protein production through acetate reduction. *Biotechnol. Prog.* **11,** 475–478 (1995).
41. Veit, A., Polen, T. & Wendisch, V. F. Global gene expression analysis of glucose overflow metabolism in *Escherichia coli* and reduction of aerobic acetate formation. *Appl. Microbiol. Biotechnol.* **74,** 406–421 (2007).
42. Galluzzi, L., Kepp, O., Vander Heiden, M. G. & Kroemer, G. Metabolic targets for cancer therapy. *Nature Rev. Drug Discov.* **12,** 829–846 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to F. J. Bruggeman, E. O'Brien, U. Sauer, M. H. Saier and members of the Hwa and Sauer laboratories for valuable comments, and J. L. Figueroa for artistic contributions to the model illustration in Box 1. This work was supported by the NIH (grant R01-GM109069) and the Simons Foundation (grant 330378). T.H. additionally acknowledges the support of M. Rössler, the Walter Haefner Foundation and the ETH Foundation. M.B. acknowledges support from SystemsX TPdF. Y.S. acknowledges support from Hong Kong Baptist University (grants FRG2/11-12/159 and SKLP-14-15-P012).

Author Contributions M.B., S.H., J.R.W. and T.H. designed the study. M.B., S.H., H.O., Z.Z. and Y.S. performed experiments. M.B., S.H. and T.H. analysed the data and developed the model. M.B., S.H., J.R.W. and T.H. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.H. (hwa@ucsd.edu).

METHODS

No statistical methods were used to predetermine sample size.

Construction of LacY and LacY^{A177V} strains (NQ1312 and NQ1313). The Ptet-lacY region of the pZE12 Ptet-lacY plasmid[43] was amplified with upstream and downstream primers including the digestion sites XhoI and BamHI, respectively, using the primers Ptet-F and lacY-R (see primers below). The resulting DNA fragment was used to replace the corresponding region of Ptet-gfp in the plasmid pZA31-gfp[44], yielding the plasmid pZA31 Ptet-lacY. This plasmid was transformed into the titratable PtsG strain NQ1243 to yield NQ1312. The same procedure was employed to generate the *lacY*^{A177V} mutant (that is, C531T), but fusion PCR was used to introduce a point mutation Val177 into the *lacY* sequence[33]. For this, two overlapping parts of the Ptet-lacY region were PCR amplified with the primers ptet-F, lacYfusion-R and lacYfusion-F, lacY-R (see primers below), in which the point substitution C531T leading to the Val177 mutation from ref. 3 was included in the primers lacYfusion-F and lacYfusion-R. These two overlapping DNA fragments were fused together by PCR using primers ptet-F and lacY-R. The resulting Ptet-lacY fragment that carries the desired mutation was inserted into pZA31, yielding pZA31-lacY^{A177V}. The resulting plasmids were transformed into the titratable PtsG strain NQ1243 to yield NQ1313.

Construction of the *flhD* and *fliA* deletion strains (NQ1388 and NQ1539). The Δ*flhD* deletion allele in strain JW1881-1 (*E. coli* Genetic Stock Center, Yale University), in which a kanamycin-resistance gene is substituted for the *flhD* gene, was transferred to the titratable PtsG strain NQ1243 after deletion of kanamycin resistance by phage P1 *vir*-mediated transduction. Similarly, the Δ*fliA* allele from strain JW1907 (KEIO collection[45]), in which a kanamycin-resistance gene is substituted for the *fliA* gene, was transferred to the titratable PtsG strain NQ1243 after deletion of kanamycin resistance by phage P1 *vir*-mediated transduction.

Primers used in this study. The following primers for producing the new genetic constructs were used. ptet-F, 5′-ACACTCGAGTCCCTATCAGTGATAGAGATTG-3′, was used for forward amplification of the Ptet sequence and included an XhoI digestion site for construction of pZE1 Ptetstab-lacZ, pZA31-lacY, pZA31-lacY^{A177V}. lacY-R, 5′-TGTGGATCCTTAAGCGACTTCATTCACCTG-3′, was used for reverse amplification of *lacY*, *lacY*^{A177V} and included a BamHI digestion site for construction of pZA31-lacY, pZA31-lacY^{A177V}. lacYfusion-F, 5′-CTCTGGCTGTGTACTCATCCTCGCCGTTTTACTCTTTTTCGCCAAAACGG-3′, was used for forward amplification of a fragment of *lacY* together with the reverse primer lacY-R. This DNA fragment was later used for fusion PCR to construct pZA31-lacY^{A177V}. lacYfusion-R, 5′-CCGTTTTGGCGAAAAAGAGTAAAACGGCGAGGATGAGTACACAGCCAGAG-3′, was used for reverse amplification of a fragment of Ptet-lacY together with the forward primer ptet-F. This DNA fragment later was used for fusion PCR to construct pZA31-lacY^{A177V}.

Bacterial culture media. Our growth media were based on the MOPS-buffered minimal medium used previously[46] with slight modifications. The base medium contains 40 mM MOPS and 4 mM tricine (adjusted to pH 7.4 with KOH), 0.1 M NaCl, 10 mM NH_4Cl, 1.32 mM KH_2PO_4, 0.523 mM $MgCl_2$, 0.276 mM Na_2SO_4, 0.1 mM $FeSO_4$ and the trace micronutrients described previously[47]. For ^{15}N-labelled media, $^{15}NH_4Cl$ was used in place of $^{14}NH_4Cl$. The concentrations of the carbon sources and various supplements used are indicated in the relevant tables.

Batch culture growth has been described in detail previously[27].

Bacterial growth in the bioreactor. To measure CO_2 production from the bacterial growth, cells were grown in a Multifors bioreactor (Infors HT). Medium (400 ml) was used in a 750-ml vessel, which has an inlet for compressed air and out outlet for the exhaust gas. The vessel is otherwise closed except during brief period of sample collection. Samples of the cell culture (for reading $A_{600\,nm}$, assaying lactose and acetate, etc) can be taken by using a syringe connected to the vessel. The air flow rate to the inlet was controlled by a mass flow controller (Cole-Parmer, 32907-67) and maintained at 400 ml min^{-1}. The outlet was connected to a BlueInOne Cell sensor unit (BlueSens) for measuring CO_2 concentration. The stir rate in the growth vessel was set as 800 r.p.m. and temperature was maintained at 37 °C.

Glucose assay. Samples (100 μl) were taken for at least eight different times during exponential growth (typically at $A_{600\,nm}$ between 0.1 and 0.6) and immediately frozen. Before the assay, samples were thawed in water and immediately centrifuged at maximum speed (13,200*g*) for 2.5 min. Supernatant (7 μl) was used to measure glucose concentrations using the Glucose Assay Kit (GAHK-20, Sigma-Aldrich). The slope of the plot of glucose concentrations versus $A_{600\,nm}$ for all replicates (multiplied with the measured growth rate) was used to determine the glucose uptake rate.

Lactose assay. To assay lactose, ∼10 μl of the collected supernatant was first digested by β-galactosidase (Sigma-Aldrich) in Z-buffer at 37 °C for 20 min. The released glucose was then assayed enzymatically by the kit commercially available (Glucose Assay Kit, GAHK20; Sigma-Aldrich). As a control, the sample was treated in the same way without β-galactosidase. Little glucose was detected in the control.

Acetate assay. Samples (200 μl) were taken for at least three different times during exponential growth (typically at $A_{600\,nm}$ between 0.1 and 0.6) and immediately frozen. Before the assay, samples were thawed in water and immediately centrifuged at maximum speed (13,200*g*) for 2.5 min. Supernatant (100 μl) were used to measure acetate concentrations using the Acetate Assay kit (10148261035, R-Biopharm). The slope of the plot of acetate concentrations versus $A_{600\,nm}$ for all replicates (multiplied with the measured growth rate) was used to determine the acetate excretion rate.

β-galactosidase assay. The assay was performed following a similar protocol as detailed in a previous study[21].

Proteomic mass spectrometry. Protein mass spectrometry samples were collected from the four bioreactor cultures, a water bath culture of equation (353) grown on glucose minimal medium, and two ^{15}N-labelled water bath cultures of NCM3722 on lactose minimal medium and NQ381 with 200 μM 3-methylbenzyl alcohol. For each of the cultures, 1.8 ml of cell culture at $A_{600\,nm} = 0.4$–0.5 during the exponential phase was collected by centrifugation. The cell pellet was re-suspended in 0.2 ml water and fast frozen on dry ice.

Sample preparation and mass spectrometry methods have been described previously[28].

Protein identification. The raw mass spectrometry data files generated by the AB SCIEX TripleTOF 5600 system were converted to Mascot generic format (mgf) files, which were submitted to the Mascot database searching engine (Matrix Sciences) against the *E. coli* SwissProt database to identify proteins. The following parameters were used in the Mascot searches: maximum of two missed trypsin cleavage, fixed carbamidomethyl modification, variable oxidation modification, peptide tolerance ±0.1 daltons (Da), MS/MS tolerance ±0.1 Da, and 1+, 2+ and 3+ peptide charge. All peptides with scores less than the identity threshold ($P = 0.05$) were discarded.

Relative protein quantification. The raw mass spectrometry data files were converted to the .mzML and .mgf formats using conversion tools provided by AB Sciex. The .mgf files were used to identify sequencing events against the Mascot database. Finally, results of the Mascot search were submitted with .mzML files to our in-house quantification software[48]. In brief, intensity is collected for each peptide over a box in retention time and m/z space that encloses the envelope for the light and heavy peaks. The data are collapsed in the retention time dimension and the light and heavy peaks are fit to a multinomial distribution (a function of the chemical formula of each peptide) using a least squares Fourier transform convolution routine[49], which yields the relative intensity of the light and heavy species. The ratio of the non-labelled to labelled peaks was obtained for each peptide in each sample.

The relative protein quantification data for each protein in each sample mixture was then obtained as a ratio by taking the median of the ratios of its peptides. No ratio (that is, no data) was obtained if there was only one peptide for the protein. The uncertainty for each ratio was defined as the two quartiles associated with the median. To filter out data with poor quality, the ratio was removed for the protein in that sample if at least one of its quartiles lay outside of 50% range of its median. Furthermore, ratios were removed for a protein in all the sample mixtures in a growth limitation if at least one of the ratios has one of its quartiles lying outside of the 100% range of the median.

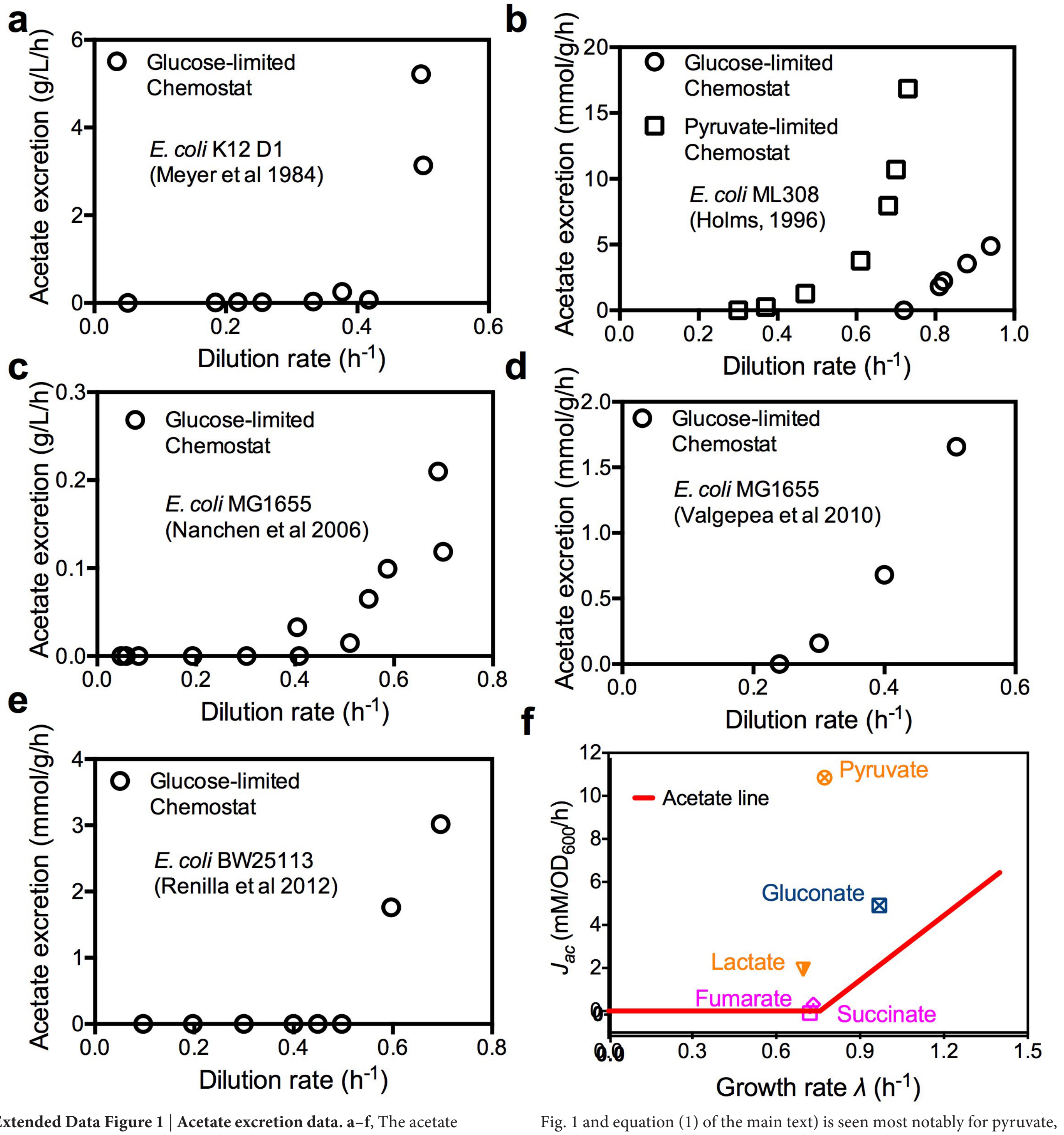
Absolute protein quantification using spectral counting data. The spectral counting data used for absolute quantitation were extracted from the Mascot search results. For our ^{15}N and ^{14}N mixture samples, only the ^{14}N spectra were counted. The absolute abundance of a protein was calculated by dividing the total number of ^{14}N spectra of all peptides for that protein by the total number of ^{14}N spectra in the sample.

Absolute quantification of LacZ protein using purified LacZ protein as standard, and determination of the converting factor between Miller Unit and proteome fraction. For the condition of the LacZ overexpression strain (NQ1389) grown on glucose medium with zero chlorotetracycline level (see source data file of Fig. 2), ^{15}N sample was prepared, that is, NQ1389 grown on glucose minimal medium with $^{15}NH_4Cl$. The sample was mixed with a known amount of purified LacZ protein (Roche Diagnostics, 10745731001), the purity of which was verified both on a SDS–PAGE gel (where a single band was observed) and by checking the spectral counts of ^{14}N peptides in the sample (where ∼99% of the ^{14}N peptides are LacZ peptides). With the highly accurate relative protein abundance between the purified ^{14}N LacZ and the ^{15}N LacZ in the sample, the proteome fraction of LacZ in the sample was determined to be 3.3% ± 0.3%. The average Miller Unit (MU) for the same condition was ∼20,550 (see source data file of Fig. 2), leading to a converting factor of 1.6% of proteome fraction for 10,000 MU.

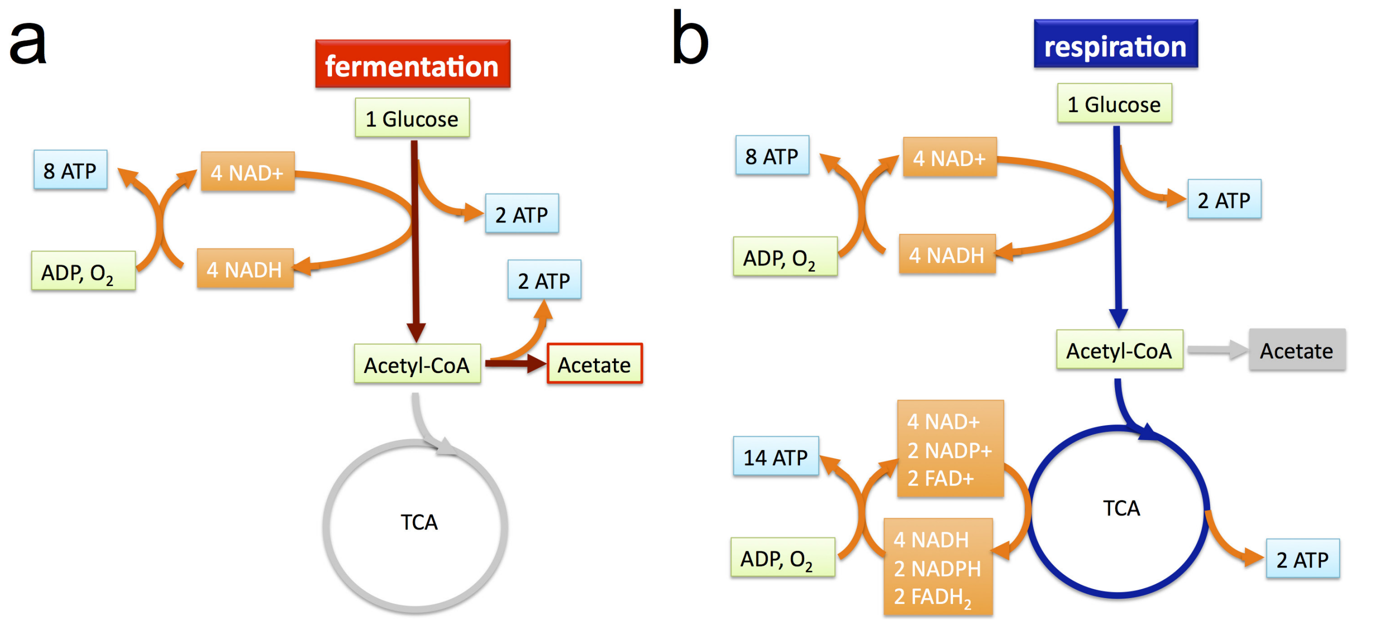
Uncertainty of individual measurements. Biological replicates show the following typical uncertainties in measured quantities: growth rate, ∼5%; acetate excretion rates, ∼15%; CO_2 evolution rate, ∼5%.

Uncertainties of linear relations. The parameters and their associated standard errors for linear relations were obtained by carrying out linear regression. Following our approach, multiple measurements over wide ranges of conditions from robust data sets revealing underlying relations between variables. The uncertainties are reported in Extended Data Tables 2 and 3, and throughout the text.

43. Kuhlman, T., Zhang, Z., Saier, M. H., Jr & Hwa, T. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **104,** 6043–6048 (2007).
44. Levine, E., Zhang, Z., Kuhlman, T. & Hwa, T. Quantitative characteristics of gene regulation by small RNA. *PLoS Biol.* **5,** e229 (2007).
45. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Sys. Biol.* **2,** 2006.0008 (2006).
46. Cayley, S., Record, M. T., Jr & Lewis, B. A. Accumulation of 3-(*N*-morpholino) propanesulfonate by osmotically stressed *Escherichia coli* K-12. *J. Bacteriol*. **171,** 3597–3602 (1989).
47. Neidhardt, F. C., Bloch, P. L. & Smith, D. F. Culture medium for enterobacteria. *J. Bacteriol.* **119,** 736–747 (1974).
48. Sperling, E., Bunner, A. E., Sykes, M. T. & Williamson, J. R. Quantitative analysis of isotope distributions in proteomic mass spectrometry using least-squares Fourier transform convolution. *Anal. Chem.* **80,** 4906–4917 (2008).
49. Rockwood, A. L. & Van Orden, S. L. Ultrahigh-speed calculation of isotope distributions. *Anal. Chem.* **68,** 2027–2030 (1996).
50. Lyons, E., Freeling, M., Kustu, S. & Inwood, W. Using genomic sequencing for classical genetics in *E. coli* K12. *PLoS ONE* **6,** e16717 (2011).
51. Soupene, E. *et al.* Physiological studies of *Escherichia coli* strain MG1655: growth defects and apparent cross-regulation of gene expression. *J. Bacteriol.* **185,** 5611–5626 (2003).
52. Brown, S. D. & Jun, S. Complete genome sequence of *Escherichia coli* NCM3722. *Genome Announc*. **3,** e00879–15 (2015).
53. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25,** 1203–1210 (1997).
54. Klumpp, S., Zhang, Z. & Hwa, T. Growth rate-dependent global effects on gene expression in bacteria. *Cell* **139,** 1366–1375 (2009).
55. Keseler, I. M. *et al.* EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* **41,** D605–D612 (2013).
56. Unden, G., Steinmetz, P. A. & Degreif-Dünnwald, P. The Aerobic and anaerobic respiratory chain of *Escherichia coli* and *Salmonella enterica*: enzymes and energetics. *EcoSal Plus* http://dx.doi.org/10.1128/ecosalplus.ESP-0005-2013 (2014).
57. Holms, H. Flux analysis and control of the central metabolic pathways in *Escherichia coli*. *FEMS Microbiol. Rev.* **19,** 85–116 (1996).
58. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25,** 25–29 (2000).
59. Macnab, R. in Escherichia coli *and* Salmonella*: Cellular and Molecular Biology* (eds Neidhardt, F. C.*et al.*) 123–145 (ASM Press, 1996).
60. McLaughlin, S. The mechanism of action of DNP on phospholipid bilayer membranes. *J. Membr. Biol.* **9,** 361–372 (1972).
61. Rhoads, D. B., Waters, F. B. & Epstein, W. Cation transport in Escherichia coli. VIII. Potassium transport mutants. *J. Gen. Physiol.* **67,** 325–341 (1976).
62. Kochanowski, K. *et al.* Functioning of a metabolic flux sensor in Escherichia coli. *Proc. Natl Acad. Sci. USA* **110,** 1130–1135 (2013).
63. Chubukov, V., Gerosa, L., Kochanowski, K. & Sauer, U. Coordination of microbial metabolism. *Nature Rev. Microbiol.* **12,** 327–340 (2014).

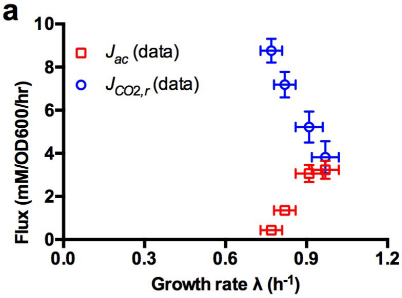


Extended Data Figure 1 | Acetate excretion data. a–f, The acetate excretion data is shown for *E. coli* cells grown in chemostat (**a**–**e**) and for cells growing on medium with non-glycolytic carbon sources (**f**). **a**, Glucose-limited chemostat data based on figure 1 from ref. 22. **b**, Glucose-limited and pyruvate–limited chemostat data from table 7 of ref. 57. **c**, Glucose-limited chemostat data based on figure 3 of ref. 23. Only data with dilution rates less than the apparent washout dilution rate are plotted here. **d**, Glucose-limited chemostat data from table 1 of ref. 15. **e**, Glucose-limited chemostat data based on figure 1 of ref. 16. **f**, *E. coli* K-12 NCM3722 was grown in minimal medium with one of five non-glycolytic carbon sources, including two gluconeogenic substrates (pyruvate and lactate), one substrate of the pentose phosphate pathway (gluconate), and two intermediates of the TCA pathway (succinate and fumarate). Deviation from the acetate line (the red line, as defined in Fig. 1 and equation (1) of the main text) is seen most notably for pyruvate, which excretes a very large amount of acetate, and to a lesser degree, also lactate and gluconate, which enter glycolysis as pyruvate. In the framework of our model, these deviations result from different proteome efficiencies of fermentation and respiration on these carbon sources. Note: acetate excretion measurement was also attempted for growth on LB. However, growth on LB is not characterized by a single exponential steady-state growth phase, as various constituents of the medium are depleted during the course of batch culture growth. Assuming exponential growth for $A_{600\,\mathrm{nm}}$ data below 0.3 and alternatively from 0.3 for 0.5 gave doubling time of 18 min and 28 min, respectively. The corresponding acetate excretion rates were 14.3 and 3.6 mM $A_{600\,\mathrm{nm}}^{-1}\,\mathrm{h}^{-1}$. These data should be regarded as semi-quantitative owing to the non-steady nature of growth on LB.



Extended Data Figure 2 | The (oxidative) fermentation and respiration pathways. a, Schematic illustration of the fermentation pathway, using glucose as an exemplary carbon source. The pathway is shown as the coloured part. One molecule of glucose is catabolized into two molecules of acetate and two molecules of CO_2 (not shown in the diagram), with four molecules of ATP generated via substrate phosphorylation and also four molecules of NADH produced. In the aerobic environment, NADH molecules can be converted into ATP molecules. The total number of ATP molecules produced per glucose molecule is therefore $4+4x$, in which the conversion factor x indicates the number of ATP molecules converted from one NADH molecule (that is, ATP:NADH $= x$:1). In the illustration, we assume that two molecules of ATP are converted from one NADH molecule; that is, $x = 2$. **b**, Schematic illustration of the respiration pathway, using glucose as an exemplary carbon source. The pathway is shown as the coloured part. One molecule of glucose is catabolized into 6 molecules of CO_2 (not shown in the diagram), producing 4 molecules of ATP, 6 molecules of NADH, 2 molecules of NADPH, and 2 molecules of $FADH_2$. Using ATP:NADH $= x$:1, ATP:NADPH $= x$:1 and ATP:$FADH_2$ $= x$:2, we have the total number of ATP molecules produced as $4+9x$ for the respiration pathway. Here in the illustration, we assumed $x = 2$. Note that the ratio of total ATP produced from respiration over total ATP produced from fermentation depends on the conversion factor x; that is, $(4+9x)/(4+4x)$. The value of this ratio ranges from 1 (for $x = 0$) to 9/4 (for $x \to \infty$).

a

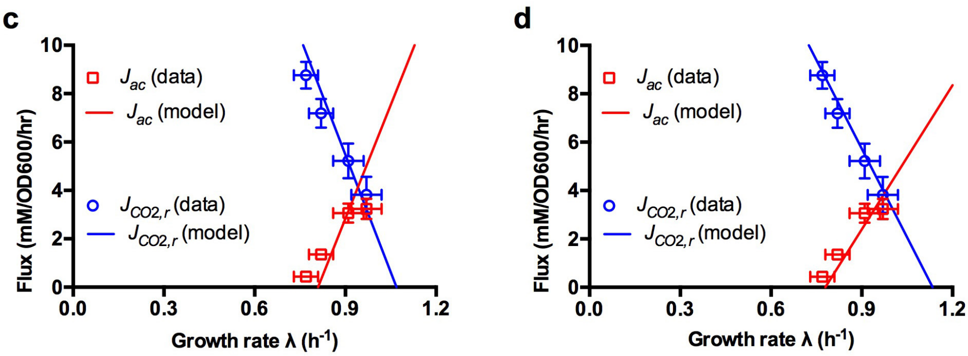


Quantity	Symbol	Unit	NQ381	NQ381	NQ381	NCM3722	Note
3MBA		μM	400	500	800	0	
Growth rate	λ	1/h	0.77	0.82	0.91	0.97	Measured
Acetate excretion rate	J_{ac}	mM/OD_{600}/h	0.44	1.35	3.06	3.24	Measured
CO_2 evolution rate	J_{CO_2}	mM/OD_{600}/h	11.96	11.50	11.56	10.57	Measured
Lactose uptake flux	J_{lac}	mM/OD_{600}/h	2.38	3.06	3.27	3.46	Measured
CO_2 flux by respiration	$J_{CO_2,r}$	mM/OD_{600}/h	8.49	6.90	4.90	3.47	Supp. Note 2

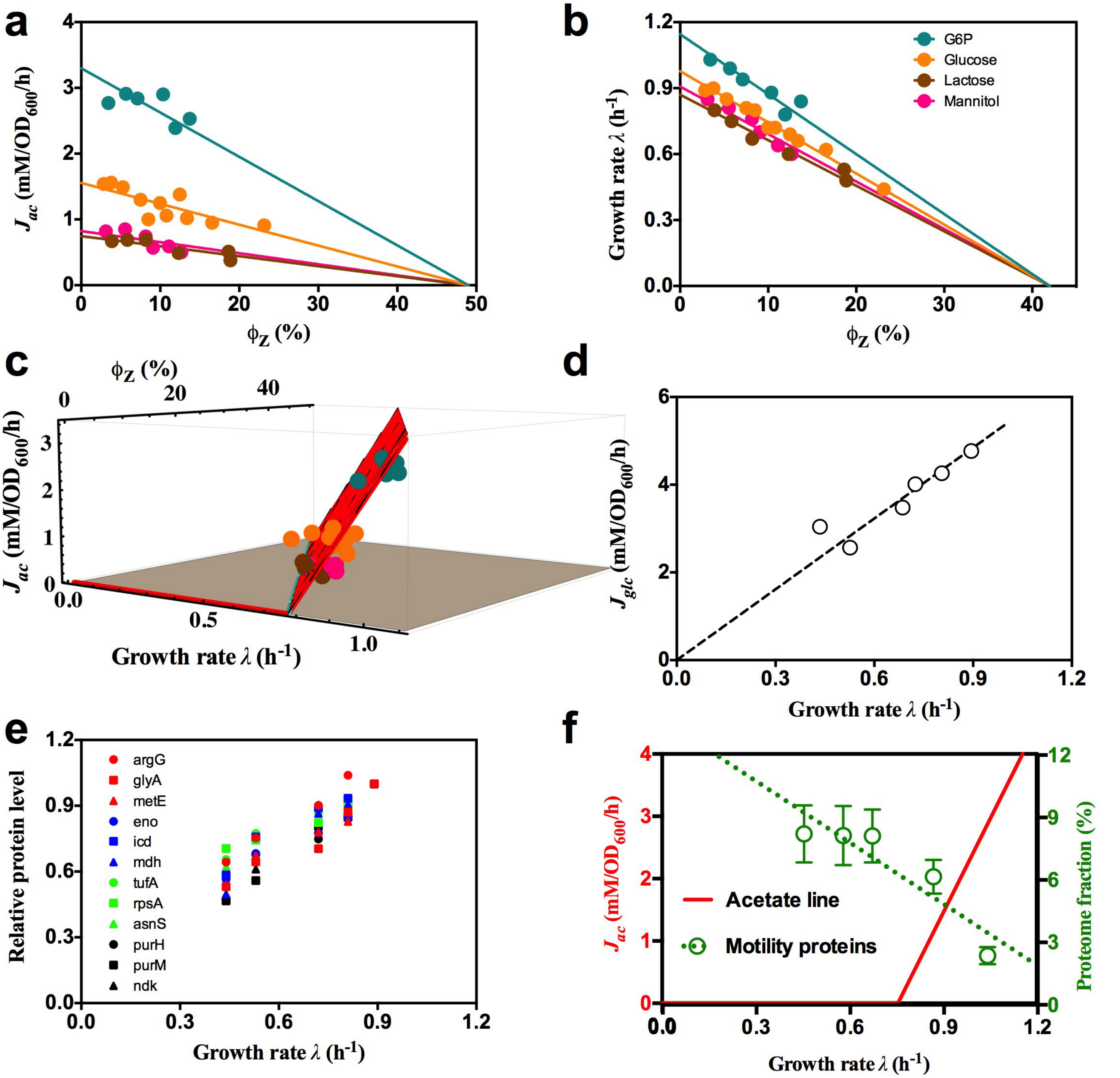
b

Quantity	Symbol or formula	NQ381	NQ381	NQ381	NCM3722
3MBA		400	500	800	0
Growth rate	λ	0.77	0.82	0.91	0.97
Carbon uptake flux	$J_{C,in} = 12 \cdot J_{lac}$	28.56	36.72	39.24	41.52
Carbon flux by fermentation	$J_{C,f} = 3 \cdot J_{ac}$	1.32	4.05	9.18	9.72
Carbon flux by respiration	$J_{C,r} = J_{CO_2,r}$	8.49	6.90	4.90	3.47
Carbon flux to biosynthesis	$J_{C,BM} = J_{C,in} - J_{C,f} - J_{C,r} = 12 \cdot J_{lac} - 3 \cdot J_{ac} - J_{CO_2,r}$	18.75	25.77	25.16	28.33
ATP flux from substrate phosphorylation by fermentation	$J_{sATP,f} = \frac{2}{3} \cdot J_{C,f} = 2 \cdot J_{ac}$	0.88	2.70	6.12	6.48
ATP flux from substrate phosphorylation by respiration	$J_{sATP,r} = \frac{2}{3} \cdot J_{C,r} = \frac{2}{3} \cdot J_{CO_2,r}$	5.66	4.60	3.26	2.32
NADH flux by fermentation	$J_{NADH,f} = \frac{2}{3} \cdot J_{C,f} = 2 \cdot J_{ac}$	0.88	2.70	6.12	6.48
NADH flux by respiration	$J_{NADH,r} = \frac{4}{3} \cdot J_{C,r} = \frac{4}{3} \cdot J_{CO_2,r}$	11.32	9.19	6.53	4.63
NADPH flux by respiration	$J_{NADPH,r} = \frac{1}{3} \cdot J_{C,r} = \frac{1}{3} \cdot J_{CO_2,r}$	2.83	2.30	1.63	1.16
$FADH_2$ flux by respiration	$J_{FADH_2,r} = \frac{1}{3} \cdot J_{C,r} = \frac{1}{3} \cdot J_{CO_2,r}$	2.83	2.30	1.63	1.16
Total ATP flux by fermentation	$J_{ATP,f} = J_{sATP,f} + 2 \cdot J_{NADH,f} = 6 \cdot J_{ac}$	2.64	8.1	18.36	19.44
Total ATP flux by respiration	$J_{ATP,r} = J_{sATP,r} + 2 \cdot J_{NADH,r} + 2 \cdot J_{NADPH,r} + 1.15 \cdot J_{FADH_2,r} = \frac{13.15}{3} \cdot J_{CO_2,r}$	37.20	30.22	21.46	15.22
Total ATP flux	$J_{ATP} = J_{ATP,f} + J_{ATP,r}$	39.84	38.32	39.82	34.66

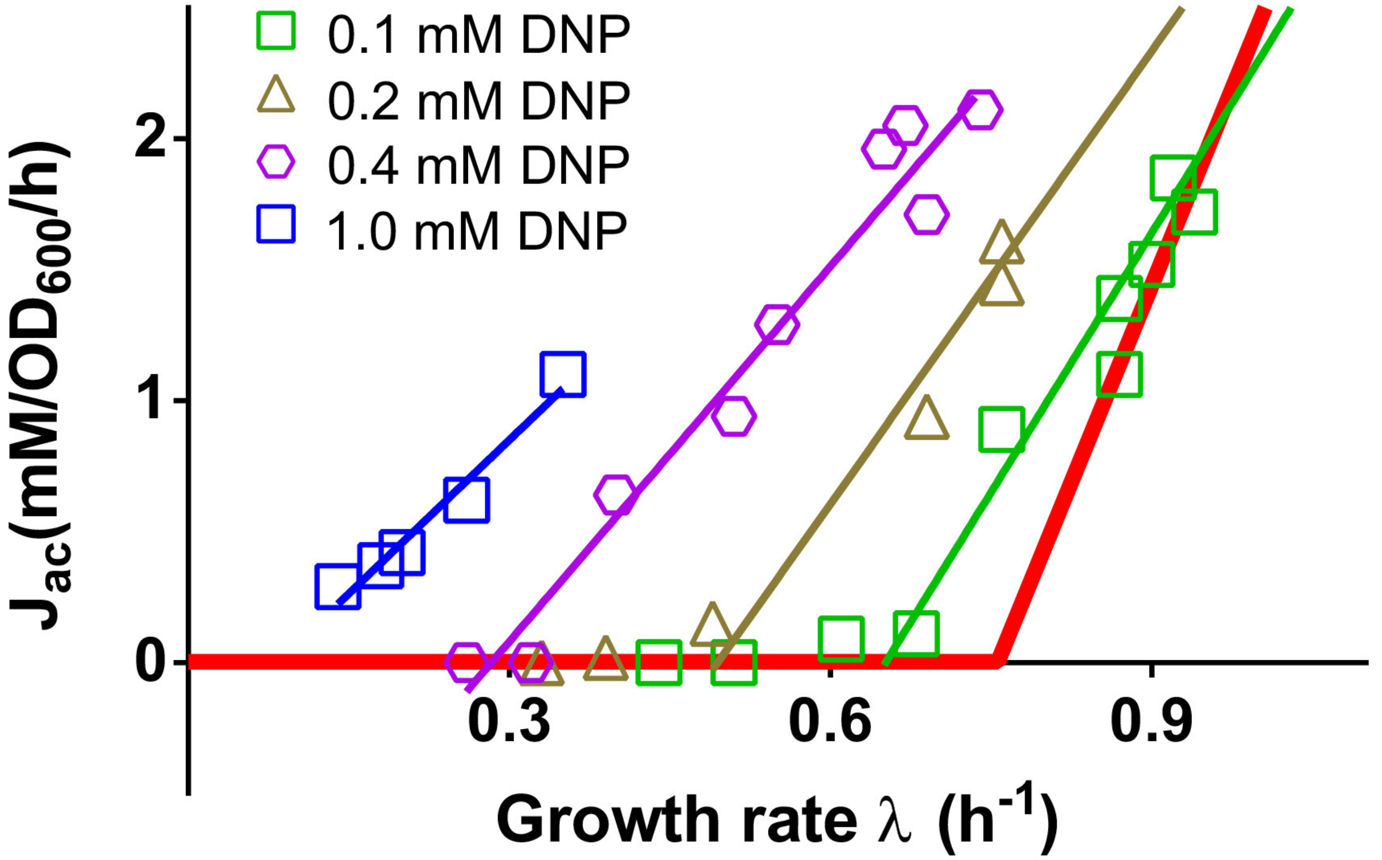
c **d**



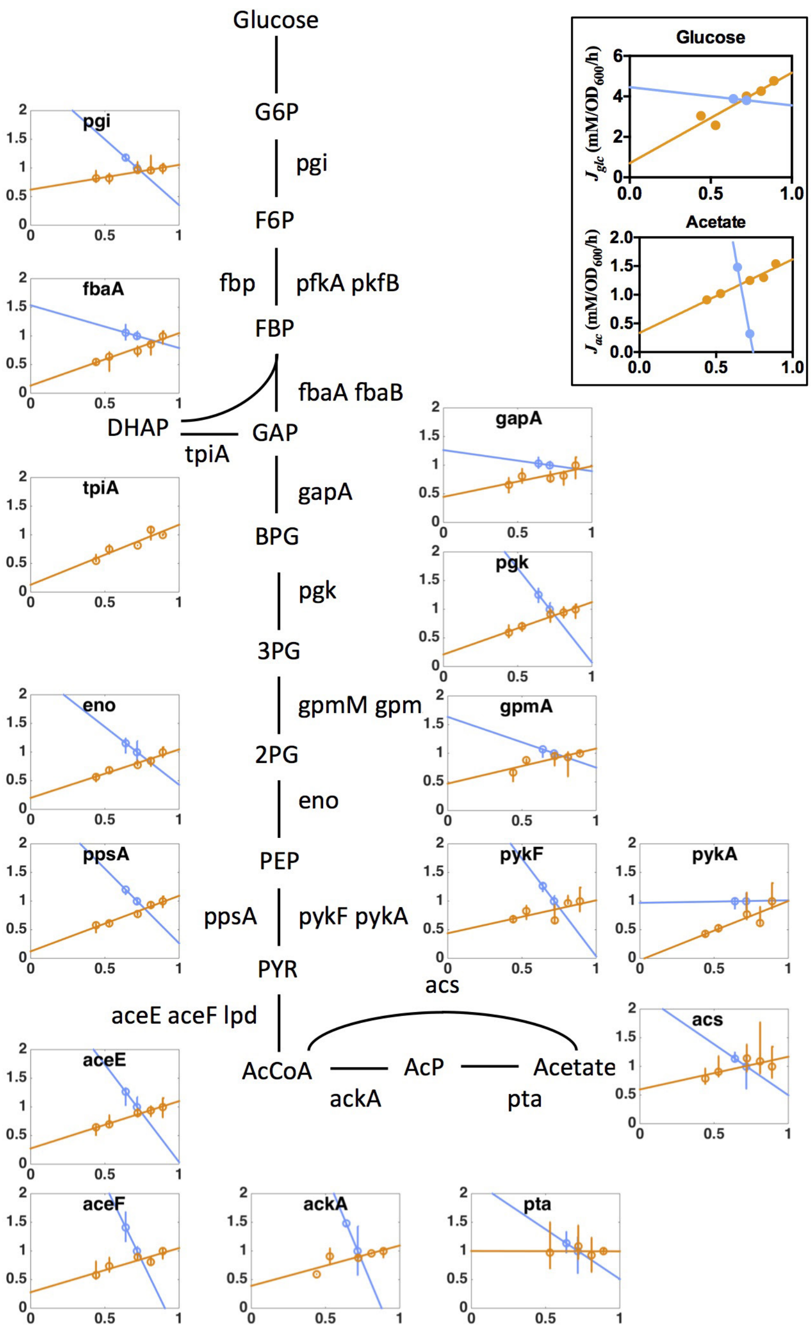
Extended Data Figure 3 | Growth-rate dependence of acetate production and CO_2 evolution in bioreactor: data and comparisons to the model. a, The rate of CO_2 evolution was determined in a bioreactor setup for wild-type and titratable LacY cells (NCM3722 and NQ381, respectively) grown in lactose minimum medium with various degrees of lactose uptake titration, and the result was used to deduce the CO_2 flux produced by respiration (blue circles). Also plotted (red squares) is the acetate excretion rate measured in the bioreactor. See Supplementary Note 2 for details of this experiment and corresponding analysis. The inducer levels, growth rate, measurements of glucose, acetate and CO_2, and the deduced CO_2 levels via respiration are shown in the table on the right. **b**, Deduced energy production fluxes from fermentation and respiration pathways, based on the measurements presented in **a**. Fluxes are in units of mM $A_{600\,\mathrm{nm}}{}^{-1}\,h^{-1}$. **c**, Comparison of model and experimental data. Using the set of parameters summarized in Extended Data Table 2, the model solution (equations (S14) and (S17)) satisfactorily describes the experimental data obtained for acetate excretion (J_{ac}) and respiratory CO_2 production ($J_{CO_2,r}$) in the bioreactor for carbon limitation. These results depend on the assumed ratios of ATP-carbon conversion. As described in Supplementary Note D1, the ratios we used in this work are ATP:NADH = 2:1, ATP:NADPH = 2:1 and ATP:$FADH_2$ = 1.15:1. Note that these conversion ratios have never been precisely measured and could be substantially overestimated[15]. However, the central results presented in this work are robust with respect to the choice of these conversion ratios. As an illustration, we show in **d** that the model results generated with a very different set of conversion ratios (ATP:NADH = 0.5:1, ATP:NADPH = 0.5:1 and ATP:$FADH_2$ = 0.5:1) even provide a slightly better description of the data. (For these conversion ratios, the energy production of the cell matches the theoretical energy demand for biomass production.) The full model calibration requires the rate of CO_2 evolution, which can only be measured in a bioreactor setup. We note a small discrepancy between acetate fluxes and growth rates obtained for cultures grown in bioreactor as compared to batch cultures, possibly caused by differences in aeration.



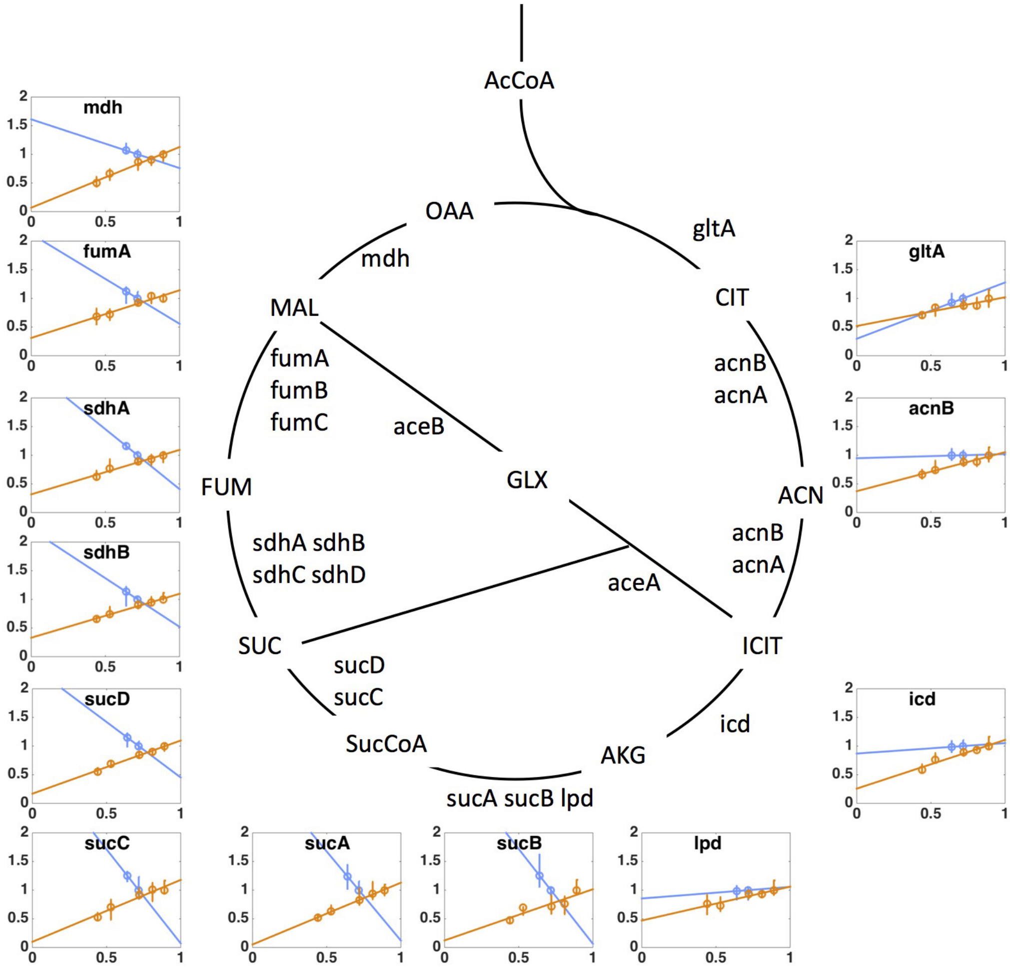
Extended Data Figure 4 | The effect of useless protein expression on acetate excretion. **a**, Acetate excretion rate by strain NQ1389 is plotted against the absolute abundance of the expressed LacZ proteins, reported as a fraction of total protein (ϕ_Z), for each of the four carbon sources described in Fig. 2a. The solid lines, depicting the linear decrease in acetate excretion, are model predictions (equation (S23) in Supplementary Information), with the lone parameter $\phi_{max} \approx 47\%$ (the x-intercept of the line) determined from the least-mean-squares fit of the data in 3D plot of Fig. 2b, by a plane anchored to the acetate line. **b**, Alternatively, ϕ_{max} can be determined from linear fits of growth rate versus ϕ_Z for the four carbon sources shown. This results in $\phi_{max} = 42\% \pm 5\%$. The solid lines in this panel are linear fits using $\phi_{max} = 42\%$. (We note that over a broad growth rate range, ϕ_{max} actually exhibits a growth-rate dependence (Dai, X. *et al.*, manuscript in preparation). Nevertheless, over the narrow growth-rate range relevant for acetate excretion, this dependence is negligible. Hence, for the purposes of our paper, we consider ϕ_{max} to be constant.) **c**, A different view of the 3D plot in Fig. 2b. **d**, Glucose uptake rate as a function of growth rate under LacZ overexpression. The circles are the data and the dashed line is the best-fit to the data passing through the origin. **e**, The relative protein levels of several representative genes, taken from amino acid synthesis (red), central metabolism (blue), protein synthesis genes (green), and nucleotide synthesis (black). As described in ref. 28, the vast majority of genes exhibited an expression pattern that is linearly proportional to the growth rate when growth is changed by increasing LacZ expression. **f**, Growth-rate dependence of motility proteins under carbon limitation. The proteome fraction data (green symbols) is from the carbon limitation series in ref. 28, in which growth rate was limited by titrating the lactose uptake for the strain NQ381. The motility proteins are proteins that are associated with the Gene Ontology (GO) term 0006810 (with GO name 'locomotion') as defined by the Gene Ontology Consortium[58]. See ref. 28 for detailed description of the experimental procedure and data processing. Note that the fraction of motility proteins increases the most in the growth range where acetate is excreted. Also note that the energy consumption by chemotaxis comprises a very minor fraction of the total energy budget, estimated to be in the order of 0.1% (ref. 59). Disabling the motility function therefore does not affect the cell's energy requirement.



Extended Data Figure 5 | Acetate excretion due to energy dissipation by DNP. DNP is a chemical known to dissipate membrane potential[60] and thus imposes energy stress on the cell. Acetate excretion rates for the titratable glucose uptake strain (NQ1243) grown in medium with glucose and different concentrations of DNP were measured for different degrees of glucose uptake. The results are qualitatively similar to the data for the leaky LacY mutant NQ1313 (Fig. 3d), except for a systematic difference in the slopes of the resulting acetate lines (thin lines of different colours). The origin of this deviation is presumably a more complex action of DNP with additional effects on the cell as compared to the leaky LacY mutant. Indeed, it is known for instance that in addition to leakage of protons, DNP also causes leakage of osmolites through the membrane[61].



Extended Data Figure 6 | The relative expression levels of glycolysis proteins under proteome perturbation (by LacZ overexpression), and energy dissipation (by expressing leaky LacY). Orange data points and linear fits result from the overexpression experiment (that is, NQ1389 grown in glucose minimal medium with different induction levels of LacZ expression), and blue data points and fits are the leaky LacY series (that is, the wild-type LacY strain NQ1312 and the leaky LacY strain NQ1313 grown in glucose minimal medium with 200 μM 3-methylbenzyl alcohol (3MBA)). The *y* axis denotes relative protein levels, which were obtained by mass spectrometry with the same reference for the two series (see Methods). The *x* axis is the growth rate (in units of h^{-1}). The different trends of protein expression for the two series show the distinct nature of the two perturbations, demonstrating that these seemingly similar predictions for the acetate line (parallel shift to slow growth as shown in Fig. 3a, d) for the two perturbations, have distinct origins and exhibit distinctly different patterns of gene expression in accordance with model predictions derived in section C3 of Supplementary Note 1 (see equations (S26) and (S36)). From the perspective of gene regulation, it is not obvious what causes the increased expression levels of glycolysis genes under energy dissipation; the transcription factor Cra in combination with the key central carbon intermediate fructose-1,6-bisphosphate (FBP) is recognized as the major regulator of glycolysis[62]. FBP relieves the repression of glycolytic enzymes expression by Cra[63]. The observed increase in the abundance of glycolytic enzymes under energy dissipation could be caused by a build-up of FBP, as energy stress limits protein polymerization. However, in this case, it is not clear what signalling pathway gives rise to the opposite responses of glycolytic enzymes to LacZ overexpression. Inset, corresponding glucose uptake and acetate excretion rates for the two perturbations presented in the main figure. Glucose uptake and acetate excretion rates decreased proportional to growth rate for LacZ overexpression (as expected from the model equations (S25), (S26) and (S29)). On the other hand, there was a marked increase in acetate excretion with energy dissipation for a roughly constant glucose uptake rate, as correctly anticipated by the model (equation (S36)). Note that the protein abundance of ACS (main panel) shows that this increase in the acetate excretion rate was not caused by a drop in ACS. Instead, the observed increase of acetate excretion, together with the parallel increase in the expression level of glycolysis and TCA enzymes, points to the coordination of glycolytic and TCA fluxes in response to energy demand.



Extended Data Figure 7 | The relative expression level of TCA proteins under proteome perturbations (by protein overexpression) and energy dissipation (by using leaky LacY). Orange data points and linear fits represent the overexpression experiment (that is, NQ1389 with different induction levels of LacZ expression), and blue data points and fits are the leaky LacY series (using strains NQ1312 and NQ1313 with wild-type and leaky LacY expression, respectively). The *y* axis denotes relative protein levels, which were obtained by mass spectrometry with the same reference for the two series (see Methods). The *x* axis is growth rate (in units of h^{-1}). The different trends of protein expression for the two series show the distinct nature of the two perturbations, demonstrating that these seemingly similar predictions on the acetate line (parallel shift to slow growth as shown in Fig. 3a, c) for the two perturbations, have distinct origins and exhibit distinctly different patterns of gene expression in accordance with model predictions derived in section C3 of Supplementary Note 1 (see equations (S27) and (S37)). From the perspective of gene regulation, the transcription factor CRP is thought to be the major regulator of TCA enzyme expression in aerobic conditions[63]. CRP-cAMP activity, which increases under carbon limitation, is known to activate the expression of most TCA enzymes. Together with the known role of CRP in upregulating the enzyme ACS which takes up acetate[15,16], CRP is considered to be a major candidate for regulating energy metabolism and acetate excretion. However, our findings that the expression of TCA enzymes increased under energy dissipation while acetate excretion also increased (Extended Data Fig. 6, inset) cannot be accounted for by known mechanisms of CRP regulation, and instead suggest an important role of additional regulators in the coordination of energy biogenesis pathways.

Extended Data Table 1 | Strains used in this study

Strain	Genotype	Description
NCM3722	wild-type *E. coli* K12 strain	parent strain for all strains used here
NQ381	*attB*::$P_{Llac\text{-}O1}$-*xylR*, *lacY*::*km-Pu-lacY*	titratable LacY
NQ636	*glpK g184t*	Glpk mutant
NQ638	*glpK a218t*	Glpk mutant
NQ640	*glpK g692a*	Glpk mutant
NQ1243	*ycaD::FRT:Ptet:xylR PptsG::kan:Pu:ptsG*	Titratable PtsG
NQ1312	*ycaD::FRT:Ptet:xylR PptsG::kan:Pu:ptsG; pZA31 Ptet-lacY*	WT LacY control for NQ1313
NQ1313	*ycaD::FRT:Ptet:xylR PptsG::kan:Pu:ptsG; Ptet-lacY*^{A177V} *on pZA31*	leaky LacY mutant
NQ1388	*ycaD::FRT:Ptet:xylR Pu:ptsG; ΔflhD-kan*	*flhD* deletion strain
NQ1389	*Ptet-tetR on pZA31; Ptetstab-lacZ on pZE1*	LacZ over-expression strain
NQ1539	*ycaD::FRT:Ptet:xylR Pu:ptsG; ΔfliA-kan*	*fliA* deletion strain
EQ353	wild-type *E. coli* MG1655 used in Li et al	obtained from Jonathan Weissman lab

Except for EQ353, all the strains used are derived from *E. coli* K-12 strain NCM3722 (refs 50–52) provided by the S. Kustu laboratory. Descriptions of the key strains used in this study are as follows. **NQ1243**: varying glucose uptake by titrating the expression of PtsG, a subunit of the glucose PTS permease. The glucose PTS permease consists of two subunits, PtsG and Crr. Strain NQ1243 was constructed by replacing the *ptsG* promoter with a titratable Pu promoter from *Pseudomonas putida*. The activity of the Pu promoter is activated by the regulator XylR upon induction by 3-methylbenzyl alcohol. Strain NQ1243 was grown in glucose minimal medium, supplemented with various 3-methylbenzyl alcohol levels (0–800 μM) to stimulate XylR and titrate the expression of PtsG.
NQ381: varying lactose uptake by titrating the expression of LacY. LacY (or lactose permease) is the primary transporter that allows *E. coli* to grow on lactose as the sole carbon source. Strain NQ381 was constructed by inserting the same titratable Pu promoter (above) between *lacZ* stop codon and *lacY* start codon. See ref. 27 details of strain construction. **NQ1389:** the titratable LacZ overexpression system. This strain carries two plasmids pZA31 and pZE1. The repressor TetR gene on the pZA31 plasmid is driven by the TetR-repressible $P_{Ltet\text{-}O1}$ promoter[53], while the *lacZ* gene on the pZE1 plasmid is driven by the modified tet-promoter (more stable with respect to spontaneous mutations). The combination of these two plasmids creates a stable, finely titratable system that can be induced via the addition of chlorotetracycline in the medium[54]. This induction system is tight, highly linear and capable of very high LacZ expression levels (with LacZ constituting up to 42% of the proteome) as seen in Fig. 2 and Extended Data Fig. 4. See ref. 28 for details of strain construction. **NQ1312 and NQ1313**: strains containing plasmids expressing LacY and LacY^{A177V}. The leaky LacY mutant (LacY^{A177V})[33] or the control wild-type LacY is each driven by the $P_{Ltet\text{-}O1}$ promoter, harboured on the pZA31 plasmid. Neither strain contains a source of the TetR repressor, hence the plasmid expression system is fully induced. Bacteria use the H^+ gradient across its inner membrane generated from the electron transport chain to produce ATP using the ATP synthase complex. The leaky LacY protein allows protons to pass through the inner membrane of the cell, thereby 'draining' the membrane potential generated by energy production pathways. This in turn leads to reduced energy efficiency or an increased energy demand on the bacterium.

Extended Data Table 2 | Model parameters calibrated from bioreactor

Parameter (Units)	Description	From literature	Measured
$\phi_{E,\max}=1-\phi_0$ (%)	maximum energy proteome fraction extrapolated to $\lambda=0$		19.0±1.2[a]
b (% hr)	energy sector growth rate dependence		12.0±1.4[b]
σ (mM/OD)	energy demand, proportionality constant with growth rate		45.7±2.8[c]
e_f (1)	carbon efficiency, fermentation	2.0[d]	
e_r (1)	carbon efficiency, respiration	4.4[e]	
ε_f (mM/OD/hr)	protein efficiency, fermentation		750±30[f]
ε_r (mM/OD/hr)	protein efficiency, respiration		390±10[g]
β (mM/OD)	carbon demand for biomass building blocks, proportionality constant with growth rate		28.5±1.3[h]
$\phi_{\max}$ (1)	maximum total proteome fraction		0.42±0.05[i]
S_{ac} (1)	stoichiometric factor for acetate from fermentation	1/3[j]	
S_{CO2} (1)	stoichiometric factor for CO_2 from respiration	1/6[k]	

[a]Determined from mass spectrometry data. Offset of linear function fitted to the total energy sector size in Supplementary Note 1, Fig. N6. Error inferred from this linear fit. See Supplementary Note 1, section D4, for details.
[b]Determined from mass spectrometry data. Slope of linear function fitted to the total energy sector size in Supplementary Note 1, Fig. N6. Error inferred from this linear fit. See Supplementary Note 1, section D4, for details.
[c]Determined from energy flux data, assuming energy demand proportional to growth rate. See Supplementary Note 1, section D3, and Fig. N5.
[d]Determined from the literature. Number of ATP produced per carbon processed in fermentation. A total of 4 ATP and 4 NADH molecules is produced per glucose molecule metabolized in fermentation (EcoCyc[55]). On the basis of ref. 56, a conversion ratio of NADH, NADPH to ATP of 2.0 was assumed. Hence, the equivalent of 2.0 ATP molecules are produced per carbon metabolized in fermentation.
[e]Determined from the literature. Number of ATP produced per carbon processed in respiration. A total of 4 ATP, 2 $FADH_2$, 2 NADPH and 8 NADH molecules is produced per glucose molecule metabolized in respiration (EcoCyc[55]). On the basis of ref. 56, a conversion ratio of NADH, NADPH to ATP of 2.0 and a conversion ratio of $FADH_2$ to ATP of 1.15 was assumed. Hence, the equivalent of 4.4 ATP are produced per carbon metabolized in respiration.
[f]Determined from mass spectrometry data. Energy flux produced per protein fraction invested in the fermentation pathway. See Supplementary Note 1, section D5, for details.
[g]Determined from mass spectrometry data. Energy flux produced per protein fraction invested in the respiration pathway. See Supplementary Note 1, section D5, for details.
[h]Determined from carbon uptake flux and carbon fluxes by the energy pathways. See Supplementary Note 1, section D2, and Fig. N4 for details.
[i]Determined by LacZ overexpression. Given by the proteome fraction occupied by LacZ, at which the growth rate vanishes. According to the average and the corresponding standard deviation of the fits of the data presented in Fig. 2, growth rate vanishes at 260,000 ± 30,000 MU, which translates into a proteome fraction of 42% ± 5%, given that 100,00 MU corresponds to 1.6% of proteome. Note that this estimate is in good agreement with the estimates in previous works[21,27,28].
[j]Here, $S_{ac} = 1/3$ simply because of the chemical reaction $6C \rightarrow 2acetate + 2CO_2$ of the fermentation pathway: the carbon uptake flux $J_{C,f}$ measured in units of the number of carbon atoms (C), is three times of the flux of acetate molecules.
[k]$S_{CO2} = 1/6$ simply because of the chemical reaction $6C \rightarrow 12CO_2$ of the respiration pathway, oxidizing all carbon atoms to CO_2.

Extended Data Table 3 | Comparison between phenomenological model predictions and empirical results

Parameter	Description	Equations	Figures	Units	Predicted	Empirical
$J_{ac}'(\lambda=0)$	LacZ overexpression, glucose, acetate excretion intercept at λ=0	S26	2a	mM/OD/hr	0.0[a]	0.17±0.11[b]
$J_{ac}'(\lambda=0)$	LacZ overexpression, G6P, acetate excretion intercept at λ=0	S26	2a	mM/OD/hr	0.0[a]	0.70±0.25[b]
$J_{ac}'(\lambda=0)$	LacZ overexpression, lactose, acetate excretion intercept at λ=0	S26	2a	mM/OD/hr	0.0[a]	0.13±0.11[b]
$J_{ac}'(\lambda=0)$	LacZ overexpression, mannitol, acetate excretion intercept at λ=0	S26	2a	mM/OD/hr	0.0[a]	-0.18±0.04[b]
s_{ac}'	slope acetate line with constant LacZ expression level of 50 000 MU (or $\phi_Z \approx 8\%$)	S30	N2a, 3a	mM/OD	10.0[c]	11.0±1.2[d]
λ_{ac}'	threshold growth rate of acetate line with constant LacZ expression level of 50 000 MU (or $\phi_Z \approx 8\%$)	S30	N2a, 3a	1/hr	0.61[e]	0.66±0.05[d]
s_{ac}'	slope acetate line with constant LacZ expression level of 100 000 MU (or $\phi_Z \approx 16\%$)	S30	N2a, 3a	mM/OD	10.0[c]	9.1±1.3[d]
λ_{ac}'	threshold growth rate of acetate line with constant LacZ expression level of 100 000 MU (or $\phi_Z \approx 16\%$)	S30	N2a, 3a	1/hr	0.47[e]	0.48±0.22[d]
λ_{ac}'	threshold growth rate of acetate line with flagella knockout	S14, S28	3b, Extended Data Fig. 4f	1/hr	0.88[f]	0.96±0.02[g]
λ_{ac}'	threshold growth rate of acetate line with 2mM chloramphenicol	S14	N2b, 3b	1/hr	0.59[h]	0.54±0.02[i]
λ_{ac}'	threshold growth rate of acetate line with 4mM chloramphenicol	S14	N2b, 3b	1/hr	0.48[h]	0.46±0.02[i]
λ_{ac}'	threshold growth rate of acetate line with 8mM chloramphenicol	S14	N2b, 3b	1/hr	0.30[h]	0.26±0.04[i]
s_{ac}'	slope of acetate line with $LacY^{A177V}$	S32	N3, 3c	mM/OD	10.0[j]	11.1±0.6[k]

[a]Equation (S26) in Supplementary Information predicts direct proportionalities between growth rate λ and acetate excretion rates J_{ac}' for LacZ overexpression. Hence, the lines should intercept the origin, with a vanishing acetate excretion rate $J_{ac}'(\lambda=0)$.

[b]Intercept $J_{ac}'(\lambda=0)$ of the least-mean-squares fit of a line to the experimental data for different levels of LacZ overexpression presented in Fig. 2a.

[c]Equation (S30) predicts a slope identical to the standard acetate line for a constant level of LacZ overexpression. The model prediction is illustrated in Supplementary Note 1, Fig. N2a, and presented as the thin red lines in Fig. 3a.

[d]For the four tested carbon sources G6P, glucose, mannitol and lactose, acetate excretion rates and β-galactosidase activities were fitted as linear functions of growth rate. These fits were then used to interpolate growth rates and acetate excretion rates for a fixed level of LacZ overexpression. This resulted in the four points for a fixed LacZ level from each of the different carbon sources presented in Fig. 3a. Resulting slopes and intercepts presented in this table are the result of least-mean-squares fits of lines to these points.

[e]Equation (S30) in Supplementary Information predicts the threshold growth rate for a fixed amount of protein overexpression, using the parameter ϕ_{max} empirically determined in this work and previous works[21,27,28] as input. The model prediction is illustrated in Supplementary Note 1 Fig. N2a and presented as the thin red lines in Fig. 3a.

[f]Using equations (S14) and (S28) of Supplementary Information, assuming the proteome sector of motility proteins decreases linearly with growth rate, vanishing at $\lambda = 1.1\ h^{-1}$ and constituting 10% of the proteome at $\lambda = \lambda_{ac}$ (compare to Extended Data Fig. 4f).

[g]Estimated from the data presented in Extended Data Fig. 4a.

[h]For chloramphenicol stress, equation (S14) in Supplementary Information predicts an increased slope s_{ac}' with an identical offset $s_{ac}'\lambda_{ac}'$, as compared to the standard acetate line. The model predictions (thin red lines, Fig. 3c) arise from using the offset of the standard acetate line given by $s_{ac}\lambda_{ac}$ as input and the slope of the modified acetate line as a fitting parameter.

[i]Threshold growth rate determined from a least-mean-squares fit of a line to the data points presented in Fig. 3c.

[j]Equation (S32) in Supplementary Information predicts an identical slope of the modified acetate line with energy dissipation as compared to the standard acetate line. The model prediction is illustrated in Supplementary Note 1, Fig. N3, and presented as the thin red line in Fig. 3d.

[k]Slope is the result of a least-mean-squares fit of a line to the data points presented in Fig. 3d.

doi:10.1038/nature16058

Warm–hot baryons comprise 5–10 per cent of filaments in the cosmic web

Dominique Eckert[1,2], Mathilde Jauzac[3,4], HuanYuan, Shan[5], Jean-Paul Kneib[5,6], Thomas Erben[7], Holger Israel[3], Eric Jullo[6], Matthias Klein[7], Richard Massey[3], Johan Richard[8] & Céline Tchernin[1]

Observations of the cosmic microwave background indicate that baryons account for 5 per cent of the Universe's total energy content[1]. In the local Universe, the census of all observed baryons falls short of this estimate by a factor of two[2,3]. Cosmological simulations indicate that the missing baryons have not condensed into virialized haloes, but reside throughout the filaments of the cosmic web (where matter density is larger than average) as a low-density plasma at temperatures of 10^5-10^7 kelvin, known as the warm–hot intergalactic medium[3–6]. There have been previous claims of the detection of warm–hot baryons along the line of sight to distant blazars[7–10] and of hot gas between interacting clusters[11–14]. These observations were, however, unable to trace the large-scale filamentary structure, or to estimate the total amount of warm–hot baryons in a representative volume of the Universe. Here we report X-ray observations of filamentary structures of gas at 10^7 kelvin associated with the galaxy cluster Abell 2744. Previous observations of this cluster[15] were unable to resolve and remove coincidental X-ray point sources. After subtracting these, we find hot gas structures that are coherent over scales of 8 megaparsecs. The filaments coincide with over-densities of galaxies and dark matter, with 5–10 per cent of their mass in baryonic gas. This gas has been heated up by the cluster's gravitational pull and is now feeding its core. Our findings strengthen evidence for a picture of the Universe in which a large fraction of the missing baryons reside in the filaments of the cosmic web.

Abell 2744 is a massive galaxy cluster (containing a total mass of $\sim 1.8 \times 10^{15}$ solar masses inside a radius of 1.3 Mpc; ref. 16) at a redshift of 0.306 (refs 17, 18). In its central regions, the cluster exhibits a complex distribution of dark and luminous matter, as inferred from X-ray and gravitational lensing analyses[16,18,19]. Spectroscopic observations indicate large variations in the line-of-sight velocity of different regions[17,18]. Together, these observations reveal that the cluster is currently experiencing a merger of at least four individual components, supporting the hypothesis that Abell 2744 may be an active node of the cosmic web.

In December 2014, we obtained a 110 ks observation of the cluster by the XMM-Newton X-ray observatory, covering the core and its surroundings out to a radius of $\sim 4h_{70}^{-1}$ Mpc, where $h_{70} = H_0/(70\ \mathrm{km\ s^{-1}\ Mpc^{-1}})$. We extracted a surface-brightness image of the observation, subtracting a model for the instrumental background and accounting for variation of the telescope efficiency across the field of view. Figure 1 shows the resulting surface-brightness image in the 0.5–1.2 keV band obtained by combining the data from the three detectors of the European Photon Imaging Camera (EPIC) on board XMM-Newton. X-ray point sources were masked and the data were adaptively smoothed to highlight the diffuse emission. The high sensitivity achieved during this observation, thanks to a minimal number of solar flares, allowed us to identify several previously unreported features. Near the virial radius of the cluster ($\sim 2h_{70}^{-1}$ Mpc) and beyond, several high-significance ($>6\sigma$) regions of diffuse emission are detected and appear to be connected to the cluster core. To confirm this connection, we extracted the X-ray emissivity profile of the cluster by masking the regions of excess emission, and compared the resulting profile with the emissivity profile in the sectors encompassing the filamentary structures (see Extended Data Fig. 1). Although the emissivity of the cluster falls below the detectable level at $\sim 2h_{70}^{-1}$ Mpc from the cluster centre, we observe significant emission in sectors extending continuously to the edge of the XMM-Newton field of view, that is, roughly at $4h_{70}^{-1}$ Mpc in projection from the core. This shows that the detected features are very extended and not caused either by the superposition of unresolved point sources or by individual group-scale haloes. These structures are not visible at higher energies (2–7 keV), in contrast to the cluster core. This suggests that the gas observed in the structures is cooler than that of the central regions.

To identify the structures detected in X-rays, we used a collection of published spectroscopic redshifts within the XMM-Newton field of view. Spectroscopic redshifts are available for 1,500 galaxies in the field[17,18]. We selected galaxies with velocities falling within $\pm 5{,}000\ \mathrm{km\ s^{-1}}$ of the cluster mean to capture the cluster and its

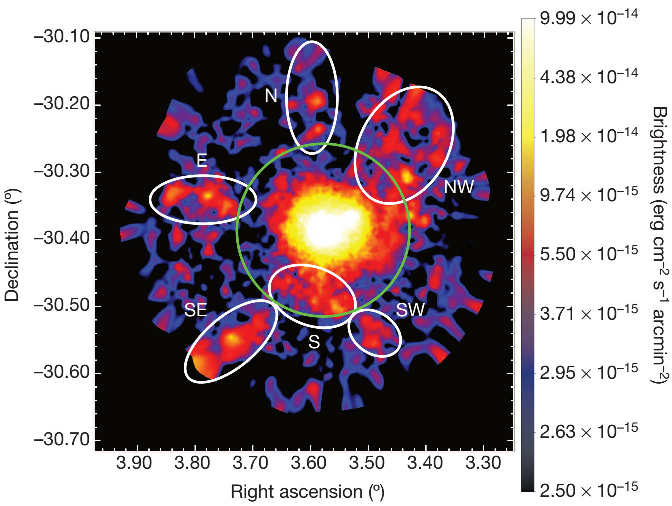


Figure 1 | Map of the hot gas in and around the galaxy cluster Abell 2744. Shown is the XMM-Newton/EPIC surface-brightness image of the galaxy cluster Abell 2744 in the 0.5–1.2 keV band. The colour bar indicates the brightness in units of erg $\mathrm{cm^{-2}\ s^{-1}\ arcmin^{-2}}$. The green circle shows the approximate location of the virial radius $R_{\mathrm{vir}} \approx 2.1h_{70}^{-1}$ Mpc. The white ellipses highlight the position of diffuse structures discovered here.

[1]Department of Astronomy, University of Geneva, Chemin d'Ecogia 16, 1290 Versoix, Switzerland. [2]INAF – IASF Milano, Via E. Bassini 15, 20133 Milan, Italy. [3]Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK. [4]Astrophysics and Cosmology Research Unit, School of Mathematical Sciences, University of KwaZulu-Natal, Durban 4041, South Africa. [5]Laboratoire d'Astrophysique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland. [6]Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, 13388 Marseille, France. [7]Argelander-Institut für Astronomie, Auf dem Hügel 71, D-53121 Bonn, Germany. [8]CRAL, Observatoire de Lyon, Université Lyon 1, 9 Avenue Ch. André, F-69561 Saint Genis Laval Cedex, France.

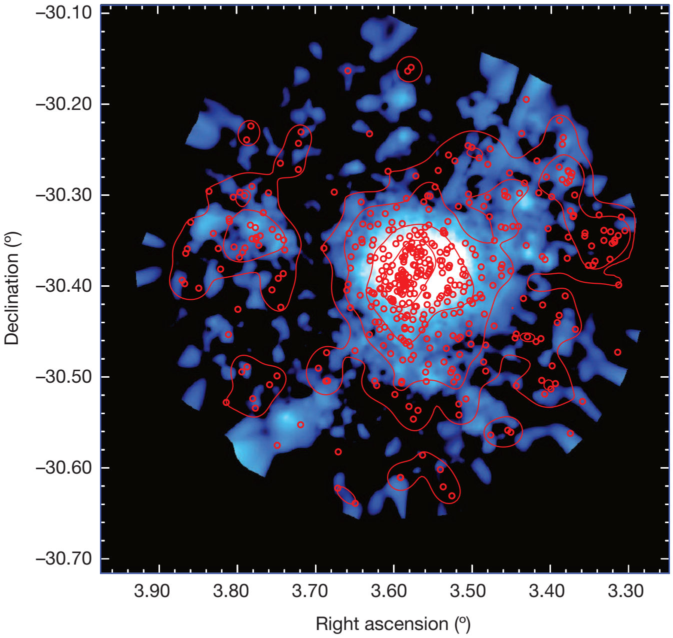


Figure 2 | Comparison between the distribution of hot gas and galaxies in the region surrounding Abell 2744. Shown is the XMM-Newton image of Abell 2744 (same data as Fig. 1); also shown are the positions of member galaxies with spectroscopic redshift within $\pm5{,}000\,\mathrm{km\,s^{-1}}$ of the cluster mean[18] (red dots); red curves show galaxy number density contours.

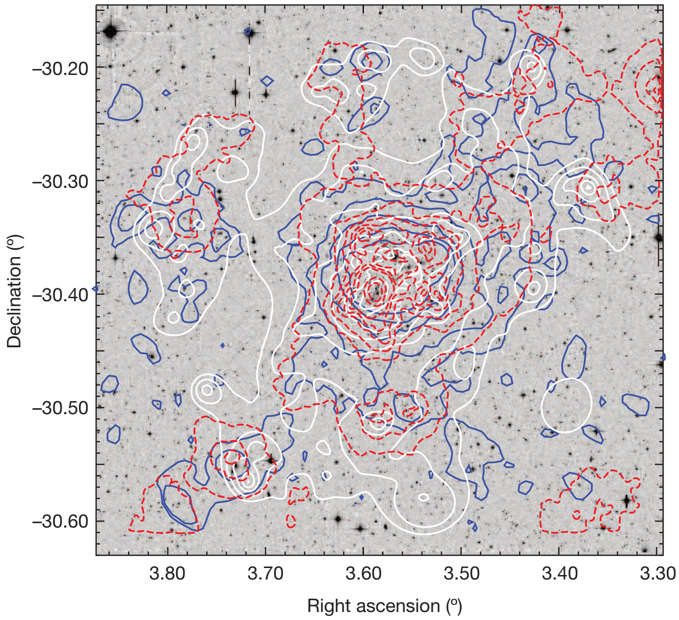


Figure 3 | Hot gas, visible light and total mass in Abell 2744. Shown is the CFHT image of Abell 2744 and the surrounding large-scale structure. The contours show X-ray isophotes (blue), mass distribution reconstructed from combined strong and weak lensing (white), and optical light (dashed red).

accretion region in their entirety. In Fig. 2 we show the XMM-Newton brightness image together with the position of selected cluster members and galaxy density contours. Concentrations of cluster galaxies are found coincident with the four hot-gas filamentary structures labelled E, S, SW and NW in Fig. 1. Conversely, structure N corresponds to a background galaxy concentration at redshift $z\approx0.45$, whereas the galaxies associated with the SE substructure exhibit a substantial velocity difference of $-8{,}000\,\mathrm{km\,s^{-1}}$ compared to the cluster core. This velocity difference corresponds to a large projected distance from the cluster, which indicates that, although it is part of the same superstructure, this system is probably not interacting with the main cluster. We therefore consider the association of the SE structure with the Abell 2744 complex as tentative and ignore it for the remainder of the analysis. As a result, we only associate structures E, S, SW and NW with the accretion flow towards Abell 2744. Structures S+SW and NW have already been identified as galaxy filaments on the basis of the galaxy distribution[15,20]. The average redshift of the galaxies in the E, S and NW structures is consistent with that of the main cluster (see Table 1), indicating that these filamentary structures are oriented close to the plane of the sky.

To map the distribution of total mass around the cluster, we measured the weak and strong gravitational lensing of background galaxies visible in wide-field optical images from ground-based telescopes and in ultra-deep Hubble Space Telescope (HST) imaging of the cluster core[21]. Our identification of cluster member galaxies utilizes a photometric galaxy catalogue based on Canada-France-Hawaii Telescope (CFHT) data in the i′ optical wavelength band and deep, archival data from the Wide-Field Imager (WFI) on the ESO 2.2-m telescope in the B, V and R bands. We selected cluster members and background galaxies using their colours in the BVRi wavelength bands[22], and used the shear signal measured from a combination of HST and CFHT images for the weak lensing analysis. We used both a simple inversion method and a combined parametric and non-parametric optimization to reconstruct the weak lensing signal. We found that all the substructures identified by XMM-Newton coincide with peaks in the matter distribution, as shown in Fig. 3. We then used the weak lensing information to infer an estimate of the mass of the structures detected in X-rays. The total mass within the identified substructures is given in Table 1. Given that dark matter dominates the total mass budget, we conclude that the structures reported here correspond to overdensities in both the baryon and dark-matter distribution.

Wide-field galaxy redshift surveys have shown that the large-scale distribution of matter in the Universe is not homogeneous[23,24]. Instead, matter tends to fall together under the action of gravity into filamentary structures, forming the cosmic web[23,25]. Galaxy clusters, the largest gravitationally-bound structures in the Universe, form at its nodes, where the matter density is the highest. We therefore associate the structures discovered here with intergalactic filaments and conclude that Abell 2744 is an active node of the cosmic web.

We estimated the plasma temperature in all the filaments highlighted in Fig. 1 by extracting their X-ray spectra and fitting them with a thin-plasma emission model. The gas in the structures has a typical density of a few times 10^{-5} particles per cm^3, corresponding to overdensities of ~200 compared to the mean baryon density[26]. Approximating

Table 1 | Properties of the filaments discovered in this study

Region	<z>	T (10^6 K)	M_{gas} ($h^{-1}\,M_\odot$)	SNR X-ray	M_{tot} ($h^{-1}\,M_\odot$)	SNR lensing	f_{gas}
E	0.308	15 ± 2	$(3.8\pm0.6)\times10^{12}$	15.4	$(7.9\pm2.8)\times10^{13}$	3.1	0.05 ± 0.02
S	0.303	16 ± 2	$(7.1\pm0.8)\times10^{12}$	22.6	$(9.5\pm2.4)\times10^{13}$	6.8	0.07 ± 0.02
SW	0.305	8^{+4}_{-2}	$(2.0\pm0.4)\times10^{12}$	9.6	$(4.8\pm1.7)\times10^{13}$	3.1	0.04 ± 0.02
NW1	0.305	25 ± 4	$(5.7\pm0.3)\times10^{12}$	25.3	$(9.5\pm2.7)\times10^{13}$	5.2	0.06 ± 0.02
NW2	0.305	19 ± 2	$(1.9\pm0.1)\times10^{13}$	25.9	$(1.2\pm0.3)\times10^{14}$	3.3	0.15 ± 0.04

X-ray and lensing properties of the regions defined in Extended Data Fig. 2. Note that because of the uncertainty in the geometry of the filaments, the provided gas mass (M_{gas}), total mass (M_{tot}) and gas fraction (f) should be considered as indicative. The masses reported here were obtained by combining strong and weak lensing. A comparison with weak-lensing-only measurements is provided in Extended Data Table 2. SNR, signal to noise ratio. $M_\odot$, solar mass.

the geometry of the filaments as segments of cylinders, we estimate the total gas mass enclosed within the filaments to be considerable ($\sim 4 \times 10^{13}$ solar masses). Given the mass within the filaments obtained from weak lensing, we estimate a gas fraction between 5% and 15% for the various substructures, depending on the adopted mass reconstruction method (see Table 1), which represents a large fraction of the Universe's baryon fraction of 15% (ref. 1). The plasma temperature is in the range $(10–20) \times 10^6$ K for the various filaments (see Table 1). This is substantially less than the virial temperature of the cluster core ($\sim 10^8$ K), which indicates that the plasma has not yet virialized within the main dark-matter halo. These gas temperatures and densities correspond to those expected for the hottest and densest parts of the warm–hot intergalactic medium (WHIM)[3,4,27,28]. Numerical simulations predict that the bulk of the gas permeating intergalactic filaments should have temperatures in the range $10^{5.5}$–$10^{6.5}$ K, but the gas in the vicinity of the cluster may have undergone substantial heating caused by adiabatic compression and shock heating. Note also that the temperatures measured here may be overestimated, given that X-ray telescopes are sensitive preferentially to the hottest phase of the expected gas distribution. Overall, these properties support the picture in which a large fraction of the Universe's baryons are located in the filaments of the cosmic web.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 May; accepted 1 October 2015.

1. Planck Collaboration XIII. Planck 2015 results. XIII. Cosmological parameters. Preprint at http://arXiv.org/abs/1502.01589 (2015).
2. Fukugita, M., Hogan, C. J. & Peebles, P. J. E. The cosmic baryon budget. *Astrophys. J.* **503,** 518–530 (1998).
3. Cen, R. & Ostriker, J. P. Where are the baryons? *Astrophys. J.* **514,** 1–6 (1999).
4. Davé, R. *et al.* Baryons in the warm-hot intergalactic medium. *Astrophys. J.* **552,** 473–483 (2001).
5. Shull, J. M., Smith, B. D. & Danforth, C. W. The baryon census in a multiphase intergalactic medium: 30% of the baryons may still be missing. *Astrophys. J.* **759,** 23 (2012).
6. Branchini, E. *et al.* Studying the warm hot intergalactic medium with gamma-ray bursts. *Astrophys. J.* **697,** 328–344 (2009).
7. Fang, T., Canizares, C. R. & Yao, Y. Confirming the detection of an intergalactic X-ray absorber toward PKS 2155–304. *Astrophys. J.* **670,** 992–999 (2007).
8. Buote, D. A. *et al.* X-ray absorption by WHIM in the Sculptor Wall. *Astrophys. J.* **695,** 1351–1356 (2009).
9. Zappacosta, L. *et al.* Studying the WHIM content of large-scale structures along the line of sight to H 2356–309. *Astrophys. J.* **717,** 74–84 (2010).
10. Nicastro, F. *et al.* Chandra view of the warm-hot intergalactic medium toward 1ES 1553+113: absorption-line detections and identifications. I. *Astrophys. J.* **769,** 90 (2013).
11. Kull, A. & Böhringer, H. Detection of filamentary X-ray structure in the core of the Shapley supercluster. *Astron. Astrophys.* **341,** 23–28 (1999).
12. Scharf, C., Donahue, M., Voit, G. M., Rosati, P. & Postman, M. Evidence for X-ray emission from a large-scale filament of galaxies? *Astrophys. J.* **528,** L73–L76 (2000).
13. Zappacosta, L. *et al.* Warm-hot intergalactic baryons revealed. *Astron. Astrophys.* **394,** 7–15 (2002).
14. Werner, N. *et al.* Detection of hot gas in the filament connecting the clusters of galaxies Abell 222 and Abell 223. *Astron. Astrophys.* **482,** L29–L33 (2008).
15. Ibaraki, Y., Ota, N., Akamatsu, H., Zhang, Y.-Y. & Finoguenov, A. Suzaku study of gas properties along filaments of A2744. *Astron. Astrophys.* **562,** A11 (2014).
16. Merten, J. *et al.* Creation of cosmic structure in the complex galaxy cluster merger Abell 2744. *Mon. Not. R. Astron. Soc.* **417,** 333–347 (2011).
17. Boschin, W., Girardi, M., Spolaor, M. & Barrena, R. Internal dynamics of the radio halo cluster Abell 2744. *Astron. Astrophys.* **449,** 461–474 (2006).
18. Owers, M. S. *et al.* The dissection of Abell 2744: a rich cluster growing through major and minor mergers. *Astrophys. J.* **728,** 27 (2011).
19. Kempner, J. C. & David, L. P. A Chandra view of the multiple merger in Abell 2744. *Mon. Not. R. Astron. Soc.* **349,** 385–392 (2004).
20. Braglia, F., Pierini, D. & Böhringer, H. Flaming, bright galaxies along the filaments of A 2744. *Astron. Astrophys.* **470,** 425–429 (2007).
21. Jauzac, M. *et al.* Hubble Frontier Fields: a high-precision strong-lensing analysis of the massive galaxy cluster Abell 2744 using 180 multiple images. *Mon. Not. R. Astron. Soc.* **452,** 1437–1446 (2015).
22. Israel, H. *et al.* The 400d Galaxy Cluster Survey weak lensing programme. I. MMT/Megacam analysis of CL0030+2618 at z=0.50. *Astron. Astrophys.* **520,** A58 (2010).
23. Bond, J. R., Kofman, L. & Pogosyan, D. How filaments of galaxies are woven into the cosmic web. *Nature* **380,** 603–606 (1996).
24. Yess, C. & Shandarin, S. F. Universality of the network and bubble topology in cosmological gravitational simulations. *Astrophys. J.* **465,** 2–13 (1996).
25. Springel, V., Frenk, C. S. & White, S. D. M. The large-scale structure of the Universe. *Nature* **440,** 1137–1144 (2006).
26. Takei, Y. *et al.* Warm-hot intergalactic medium associated with the Coma Cluster. *Astrophys. J.* **655,** 831–842 (2007).
27. Dolag, K., Meneghetti, M., Moscardini, L., Rasia, E. & Bonaldi, A. Simulating the physical properties of dark matter and gas inside the cosmic web. *Mon. Not. R. Astron. Soc.* **370,** 656–672 (2006).
28. Gheller, C., Vazza, F., Favre, J. & Brüggen, M. Properties of cosmological filaments extracted from Eulerian simulations. *Mon. Not. R. Astron. Soc.* **453,** 1164–1185 (2015).

Acknowledgements Work reported here is based on observations obtained with XMM-Newton, an ESA science mission with instruments and contributions directly funded by ESA Member States and NASA. D.E. thanks F. Vazza, S. Paltani and S. Molendi for discussions. We thank H. Ebeling, M. Limousin, B. Clément, H. Atek, D. Harvey, E. Egami, M. Rexroth and P. Natarajan for help with writing the XMM-Newton proposal. M.J., H.I. and R.M. acknowledge support from the UK Science and Technology Facilities Council (grant numbers ST/L00075X/1, ST/H005234/1), the Leverhulme trust (grant number PLP-2011-003) and the Royal Society. J.-P.K. acknowledges support from the ERC advanced grant LIDA and from CNRS. H.Y.S. acknowledges support by a Marie Curie International Incoming Fellowship within the 7th European Community Framework Programme, and NSFC of China under grant 11103011. T.E. was supported by the Deutsche Forschungsgemeinschaft through the Transregional Collaborative Research Centre TR 33 'The Dark Universe'. E.J. was supported by CNES. J.R. acknowledges support from the ERC starting grant CALENDS.

Author Contributions D.E.: lead author, X-ray analysis. M.J.: weak and strong lensing analysis. H.Y.S.: CFHT weak lensing analysis. J.-P.K.: principal investigator of the XMM-Newton observation, strong and weak lensing analysis and identification of the red cluster sequence in the photometric data. T.E.: WFI and CFHT data reduction. H.I.: WFI and CFHT data reduction. E.J.: weak and strong lensing modelling techniques. M.K.: WFI and CFHT data reduction. R.M.: weak lensing analysis. J.R.: strong lensing analysis. C.T.: X-ray analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.E. (Dominique.Eckert@unige.ch).

METHODS

Imaging X-ray analysis. Abell 2744 was observed by XMM-Newton in late 2014 for a total observing time of 110 ks (PI, J.-P.K.; OBSID, 074385). At the redshift of Abell 2744 (0.306), the size of the XMM-Newton field of view corresponds to $8h_{70}{}^{-1}$ Mpc. We processed the data using the XMM-Newton Scientific Analysis System (XMMSAS) v14.0. We excluded flaring periods from the event files by creating a light curve for each instrument separately and filtering out the time periods for which the observed count rate exceeded the mean by more than 2σ. The observation was very mildly affected by soft-proton flares, allowing us to reach a flare-free observing time of 96 ks, 97 ks and 87 ks for EPIC detectors MOS1, MOS2 and pn, respectively.

We extracted raw images in the 0.5–1.2 keV band for all three EPIC detectors using the Extended Source Analysis Software (ESAS) package[29]. This energy band maximizes the source-to-background ratio and avoids the bright Al and Si background emission lines, while maintaining a large effective area since the collecting power of the XMM-Newton telescopes peaks at 1 keV. Exposure maps for each instrument were created, taking into account the variations of the vignetting across the field of view. A model image of the non X-ray background (NXB) was computed using a collection of closed-filter observations, and was adjusted to each individual observation by comparing the count rates in the corner of the field of view. X-ray point sources were detected using the XMMSAS tool ewavelet and masked during the analysis. Additionally, we used the existing Chandra observations of the cluster[18,19] to detect point sources down to fainter X-ray fluxes ($\sim5\times10^{-16}$ erg cm^{-2} s^{-1}) and mask the corresponding areas. Such a flux threshold for point-source removal corresponds to a resolved fraction of 80% of the cosmic X-ray background[30], which is associated with a cosmic variance of about 5%. This ensures that the extended features reported here are indeed caused by diffuse emission.

We computed surface-brightness images by subtracting the NXB from the raw images and dividing them by the exposure maps. To maximize the signal-to-noise ratio (SNR), we then combined the surface-brightness images of the three EPIC detectors by weighting each detector by its relative effective area. The resulting image was then adaptively smoothed using the XMMSAS tool asmooth, requiring an SNR of 5 for all features above the local background. The total XMM-Newton/EPIC image of Abell 2744 is shown in Fig. 1.

To confirm the presence of the filamentary structures shown in Fig. 1, we compared the surface brightness of the regions inside and outside the filaments. We used the PROFFIT code[31] to extract the surface brightness profile from the surface-brightness peak by masking the sectors corresponding with the filaments, and we compared the masked profile with the surface brightness profile in the direction of the filaments, that is, in the sectors including the filaments (position angles 10°–70°, 150°–180° and 260°–300° for the NW, E and S filaments, respectively, where 0° is the W direction; see Extended Data Fig. 2). In Extended Data Fig. 1 we show the corresponding surface-brightness profiles. When masking the filaments, no statistically significant cluster emission is detected beyond 7 arcmin ($\sim2h_{70}{}^{-1}$ Mpc); in the direction of the filaments, a flat surface brightness is observed out to the edge of the field of view ($\sim4h_{70}{}^{-1}$ Mpc). The small variations in the amplitude of the surface-brightness profiles indicates that the emission is due to filamentary structures rather than to a collection of infalling clumps. The excess emission produced by the filaments has already been noted in Suzaku observations of the cluster[15]; the poor angular resolution and narrow field of view of Suzaku were however insufficient to separate the filaments from the field and resolve point sources.

For comparison, we extracted radial profiles of galaxy density from spectroscopically-confirmed members[18] in exactly the same sectors. The resulting profiles are shown in Extended Data Fig. 3. We find that beyond the cluster's virial radius the galaxy density is consistently larger in the regions containing the filaments compared to the perpendicular directions, which highlights the association between the structures detected in X-rays and the local galaxy distribution.

Spectral X-ray analysis. We performed a spectral analysis of the structures highlighted in Fig. 1. We defined elliptical regions following the X-ray isophotes as closely as possible. In Extended Data Fig. 2 we show the regions used to derive the spectral properties of the filaments. Since the surface brightness of these regions barely exceeds the background level, a detailed modelling of all the various background components is necessary to obtain reliable measurements of the relevant parameters. We adopted the following approach to model the various spectral components[32].

The source. We modelled the diffuse emission in each region using the thin-plasma emission code APEC[33], leaving the temperature and normalization as free parameters. The metal abundance Z was fixed to $0.2Z_{\odot}$ (ref. 34). This component is absorbed by the Galactic column density, which we fixed to the 21-cm value ($N_{\rm H}=1.5\times10^{20}$ cm^{-2}; ref. 35).

The non-X-ray background (NXB). We used closed-filter observations to estimate the spectrum of the NXB component in each region[29]. Instead of subtracting the NXB, we modelled it using a phenomenological model and included it as an additive component in the spectral fitting. This method has the advantage of retaining the statistical properties of the original spectrum. We left the normalization of the NXB component free to vary during the fitting procedure, which allows us to take variations of the NXB level into account. The normalization of the prominent background lines was also left free. Since the observation was very weakly contaminated by soft proton flares, the residual soft proton component can be neglected.

The sky background components. We used 4 offset regions where no cluster emission is detected (see Extended Data Fig. 2) to measure the sky background components in the field of Abell 2744. We modelled the sky background using a three-component model: (i) a power law with photon index fixed to 1.46 to model the cosmic X-ray background (CXB); (ii) a thermal component at a free temperature to estimate the Galactic halo emission; and (iii) an unabsorbed thermal component at 0.11 keV for the local hot bubble. The best-fit spectrum for the Offset 1 region is shown in the top-left panel of Extended Data Fig. 4. In Extended Data Table 1 we show the best-fit parameters for our sky background model in the four offset regions. The variation of the parameters from one region to another allows us to estimate the systematic uncertainties associated with the variation of the sky background across the field of view. The main sky component (the CXB) typically varies by $\pm10\%$ across the field. Slightly larger variations ($\sim20\%$) are observed for the foreground components, although it must be noted that the normalizations of the Galactic halo and local bubble components are correlated. The overall values of these parameters agree well with previous measurements of the CXB[36] and the foregrounds[37].

We note that because of strong Galactic absorption in the far-ultraviolet band and falling effective area in this wavelength range, XMM-Newton is sensitive predominantly to the hottest phase of the gas ($T>10^{6.5}$ K). To test the sensitivity of our observations to cooler plasma, we assumed a differential emission-measure model including gas temperatures in the range $10^{5.5}$–10^{7} K and simulated an XMM-Newton spectrum at the same depth as our observation. The resulting spectrum can be well fitted with a single-temperature model at $T=10^{6.8}$ K. This indicates that the temperatures measured here may be substantially overestimated if the plasma is multiphase.

In Extended Data Fig. 4 we show the observed spectra for the five regions defined in Extended Data Fig. 2 together with their best-fit model. Since it is the brightest and most extended, the NW filament was split into two regions (labelled NW1 and NW2) to study the variation of the spectral parameters along a single filament. The resulting parameters are provided in Table 1. To estimate the gas mass within each filamentary structure, we modelled the emission region as a cylinder with length and diameter given by the major and minor axes of the defined ellipses, respectively. We converted the measured normalization into an emission measure, and computed the average gas density assuming constant density in each structure. We estimated the gas mass by integrating the resulting gas density over the volume (see Table 1). We note that given the large uncertainties in the 3D geometry of the filaments, the recovered gas densities and masses should be considered as indicative. Indeed, we tested the effect of adopting different geometries (spheres, ellipsoids) on the recovered gas mass and gas density, and found that the results obtained with the various geometries vary by $\sim30\%$.

To assess the level of systematic uncertainties in our spectral measurements, we used the spectrum of the SW region, as it is the weakest and thus is the most prone to systematic uncertainties, and let the various sky background and NXB parameters vary within their allowed ranges. We then applied a Markov chain Monte Carlo (MCMC) algorithm to sample the likelihood distribution. The posterior distribution for the measured parameters are then marginalized over the systematic uncertainties associated with the variation of the background components. Through this approach, we found a typical systematic uncertainty of $\sim20\%$ on the gas temperature and $<5\%$ on the emission measure. These values provide an upper limit to the level of systematic uncertainties in the other regions since the intensity of the source relative to the background is higher than for the analysis carried out here.

Analysis of ESO and CFHT optical data. We used the colours of galaxies in archival optical imaging of the Abell 2744 field to identify members of the cluster and its associated filaments. We constructed a photometric catalogue from observations obtained in the B,V and R filters using the WFI instrument at the ESO 2.2-m telescope at La Silla Observatory, combined with i-band data obtained with MegaCam/MegaPrime at the CFHT. For the WFI BVR filters, we were able to use existing co-added images (B, 9,200 s; V, 8,700 s, R, 21,000 s) from a weak lensing follow-up of clusters detected in the Sunyaev–Zeldovich (SZ) effect. Observations spanning three campaigns between September 2000 and October 2011 were bias-subtracted and flat-fielded using the THELI processing pipeline[38,39]. THELI also includes astrometric, relative and absolute photometric calibration. Finally, the CFHT i-band data obtained in July 2009 were reduced using the CFHT-specific THELI

adaptation developed and applied for the CFHTLenS project[40]. For all filters, the co-added images were post-processed, and saturated stars and otherwise unreliable image areas were masked out[41]. Source catalogues were distilled from the co-added images using the weak lensing pipeline from ref. 22. Because of the different field-of-view of the cameras involved ($34' \times 34'$ for WFI versus $60' \times 60'$ for CFHT MegaCam), it proved useful to adopt the following strategy: we measured source photometry in all three WFI passbands in one go, making use of the double detection mode in SEXTRACTOR[42], with the deep R-band data as the detection image. In order to obtain consistent magnitudes, photometric quantities were measured after having matched the seeing in the other filters to the poorest seeing among them. A separate detection run was performed for the CFHT i′-data. The output catalogues were merged, identifying as the same object sources detected in WFI and CFHT within 0.5 arcsec of each other, yielding a common photometric catalogue containing 37 WFI galaxies per square arcmin. Objects were categorized as stars or galaxies based on their apparent size and magnitude.

Lensing analysis of HST and CFHT data. *Lensing constraints from the HST field of view.* The strong lensing constraints used to model the inner core of Abell 2744 consist of a set of 51 multiply-imaged systems (159 images[15]). The weak lensing catalogue for the HST field of view was built following the methods described in ref. 43, and the details of the Abell 2744 weak-lensing catalogue will be given elsewhere (M.J. *et al.*, manuscript in preparation). Here we give a brief summary of the different steps.

The weak lensing analysis is based on shape measurements in the Advanced Camera for Surveys (ACS)/F814W band. Following a method developed for the analysis of data obtained for the COSMOS survey[44], the SEXTRACTOR photometry package[42] was used for the detection of the sources. The resulting catalogue was then cleaned by removing spurious sources, duplicate detections, and any sources in the vicinity of stars or saturated pixels. Finally, to overcome the pattern-dependent correlations introduced by the drizzling process between neighbouring pixels, we simply scaled up the noise level in each pixel[44] by the same constant $FA \approx 0.316$ (ref. 45).

Since only galaxies behind the cluster are gravitationally lensed, the presence of cluster members dilutes the observed shear and reduces the statistical significance of all quantities derived from it. Therefore, the identification and removal of the contaminating unlensed galaxies is crucial. Thanks to the HST data in three bands (F814W, F606W and F435W), we identified the foreground galaxies and cluster members using a colour–colour diagram[21]. The measure of galaxy shapes was done using the Rhodes-Refregier-Groth (RRG) method[46], adapted to multi-epoch images like the one coming from the HSTFF data of Abell 2744[47]. Finally, galaxies with ill-determined shape parameters were excluded, since these galaxies do not contribute substantially to the shear signal[21,43].

Lensing constraints from the CFHT field of view. We employed the popular Kaiser-Squires-Broadhurst (KSB) method for galaxy shear measurement[48]. We modelled the observed galaxy shape as a convolution of the (sheared) galaxy with the point spread function (PSF), which is itself modelled as a circular profile convolved with a small anisotropy. For the PSF modelling, we identified stars in the size–magnitude and $\mu_{\max}$–magnitude planes chip by chip[49], where $\mu_{\max}$ is the peak surface brightness. We then measured the Gaussian-weighted shape moments of the stars, and constructed their ellipticity. In addition to cuts in $\mu_{\max}$ and magnitude, we also excluded noisy outliers with SNR <100 or absolute ellipticity more than 2σ away from the mean local value, and we iteratively removed objects very different from neighbouring stars. Having obtained our clean sample of stars, a second-order polynomial model in x and y was used to model the PSF across the field of view. The ellipticity of the PSF changes from its core to its wings. We measured the PSF shape using weight functions of different sizes and, when correcting each galaxy, used the weight function of the same size to measure the shapes of both the PSF and the galaxies. Background galaxies were selected with the magnitude cuts $20 < i' < 26$, size cuts $1.15 r_{\mathrm{PSF}} < r_{\mathrm{h}} < 10$ pixels (where r_{h} is the half-light radius and r_{PSF} is the size of the largest star), SNR > 10 and SEXTRACTOR flag FLAGS $= 0$. After masking and catalogue cuts, the galaxy number density is ~ 10 galaxies per square arcmin. We then measured the shapes of all the selected galaxies. Our implementation of KSB is based on the KSBf90 pipeline[49]. Details of the calibration and systematic effects are shown and discussed elsewhere[49]. If the PSF anisotropy is small, the shear γ can be recovered to first order from the observed ellipticity e^{obs} of each galaxy via

$$\gamma = P_\gamma^{-1}\left(e^{\mathrm{obs}} - \frac{P^{\mathrm{sm}}}{P^{\mathrm{sm}*}}\right)e^*$$

where asterisks indicate quantities that should be measured from the PSF model interpolated to the position of the galaxy, P^{sm} is the smear polarizability, and P_γ is the correction to the shear polarizability that includes smearing with the isotropic component of the PSF. The ellipticities were constructed from a combination of each object's weighted quadrupole moments, and the other quantities involve higher-order shape moments. All definitions are taken from ref. 50. Note that we approximate the matrix P_γ by a scalar equal to half its trace. Since measurements of Tr P_γ from individual galaxies are noisy, we fit it as a function of galaxy size and magnitude, which are more robustly observable galaxy properties[49].

The weight of the shear contribution from each galaxy is defined as

$$w = \frac{P_\gamma^2}{\sigma_0^2 P_\gamma^2 + \sigma_{e,i}^2}$$

where $\sigma_{e,i}$ is the error for an individual galaxy obtained via the formula in Appendix A of ref. 51, and $\sigma_0 \approx 0.3$ is the dispersion of the intrinsic ellipticities of galaxies. With the help of the shear catalogue, we then estimated the total mass within the filaments. As the weak lensing effect is not very sensitive to the mass profile, we assumed a dual pseudo isothermal elliptical (dPIE) profile centred on the X-ray position to measure the total mass of the filament candidates using the parametric model-fitting algorithm LENSTOOL[52]. As the weak lensing effect is not very sensitive to the mass profile, we also tested the accuracy of the derived masses by fitting again the shear profile with an elliptical Navarro-Frenk-White (NFW) profile with a concentration $c = 1$. The measured masses are consistent within the uncertainties.

Lensing mass model. The mass model built for this analysis used strong and weak lensing constraints, combining parametric and non-parametric approaches to model the global mass distribution[53]. The details of the mass modelling will be given elsewhere (M.J. *et al.*, manuscript in preparation). We kept the parametric model built for the strong lensing analysis of Abell 2744 fixed to their best-fit values, and we modelled the surrounding mass distribution using a multi-scale grid drawn from a prior light distribution of the cluster using the WFI multi-band photometric catalogue. The nodes of the grid model were parameterized using Radial Basis Functions (RBFs[54]). This allowed us to appropriately weight the strong lensing constraints without taking them twice into account[53].

The strong lensing parametric model was composed of two cluster-scale haloes. The multi-scale grid was composed of 10,282 RBFs, for which only the amplitude was left free while fitting. To the 733 cluster members identified in the HST fields of view, we added 1,457 cluster members identified using a standard colour–magnitude selection using B, V and R bands coming from WFI observations to identify the red-sequence galaxies of the cluster. Galaxy-scale haloes were modelled as RBFs, using dPIE potentials. The resulting mass map is shown by the white contours in Fig. 3.

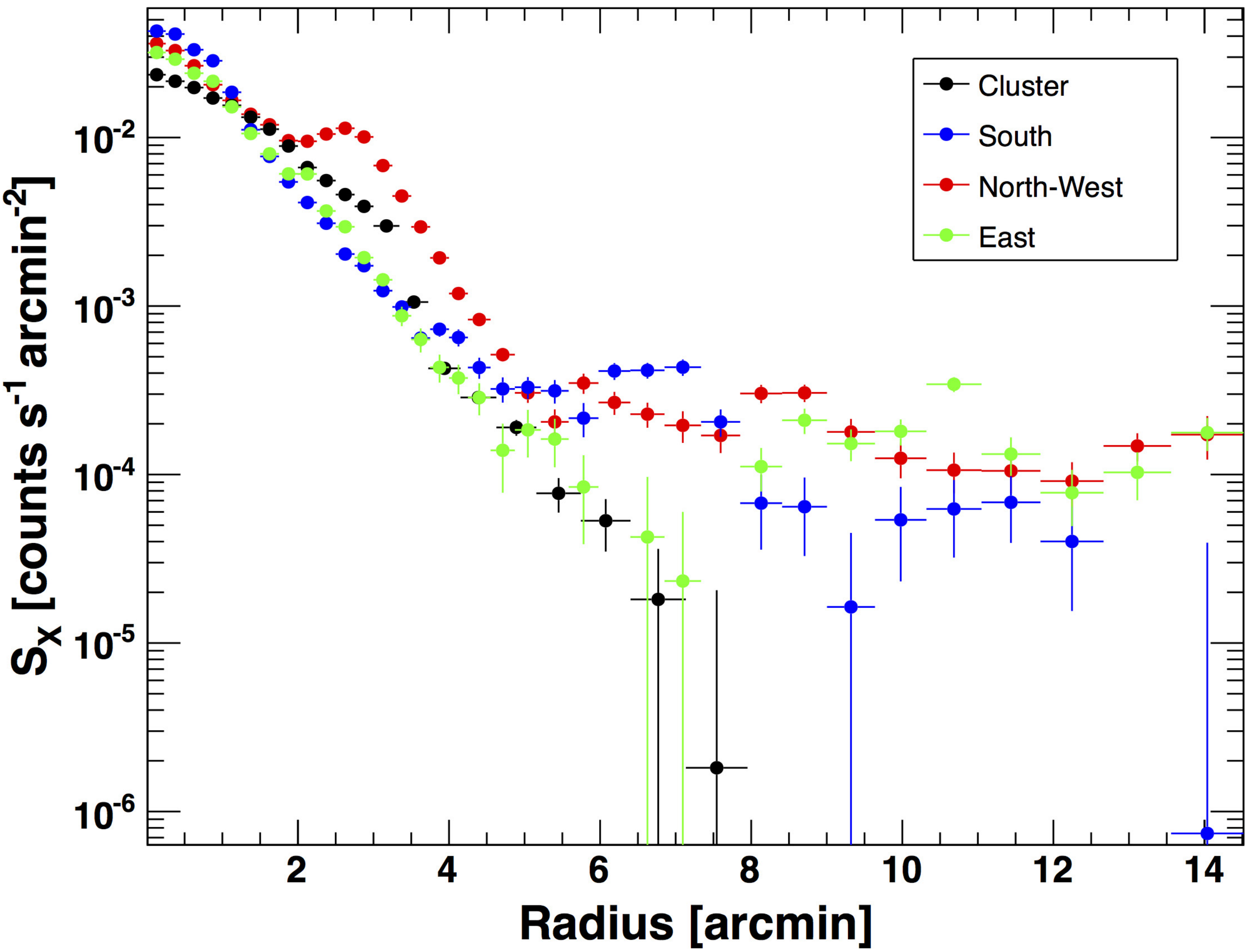
We sampled the parameter space in LENSTOOL using the Bayessys Library implemented in LENSTOOL[52]. The objective function is a standard likelihood function in which noise is assumed to be Gaussian. LENSTOOL returns a large number of MCMC samples, from which we estimate mean values and uncertainties in the mass density field. In Extended Data Fig. 5 we show the radial surface mass density profile for the cluster average compared to the sectors encompassing the filaments (same as for Extended Data Fig. 1). An excess lensing signal is observed in the direction of the filaments compared to the radial average. The masses obtained using this technique are given in Table 1. In Extended Data Table 2 we show the masses and SNRs obtained using this method (hybrid LENSTOOL) and the direct inversion method described above (KSB) for the various filaments. The results of the two methods agree within the uncertainties. The differences observed between one method and the other allow us to quantify the level of systematic uncertainties associated with the lensing reconstruction using the existing data.

Sample size. No statistical methods were used to predetermine sample size.

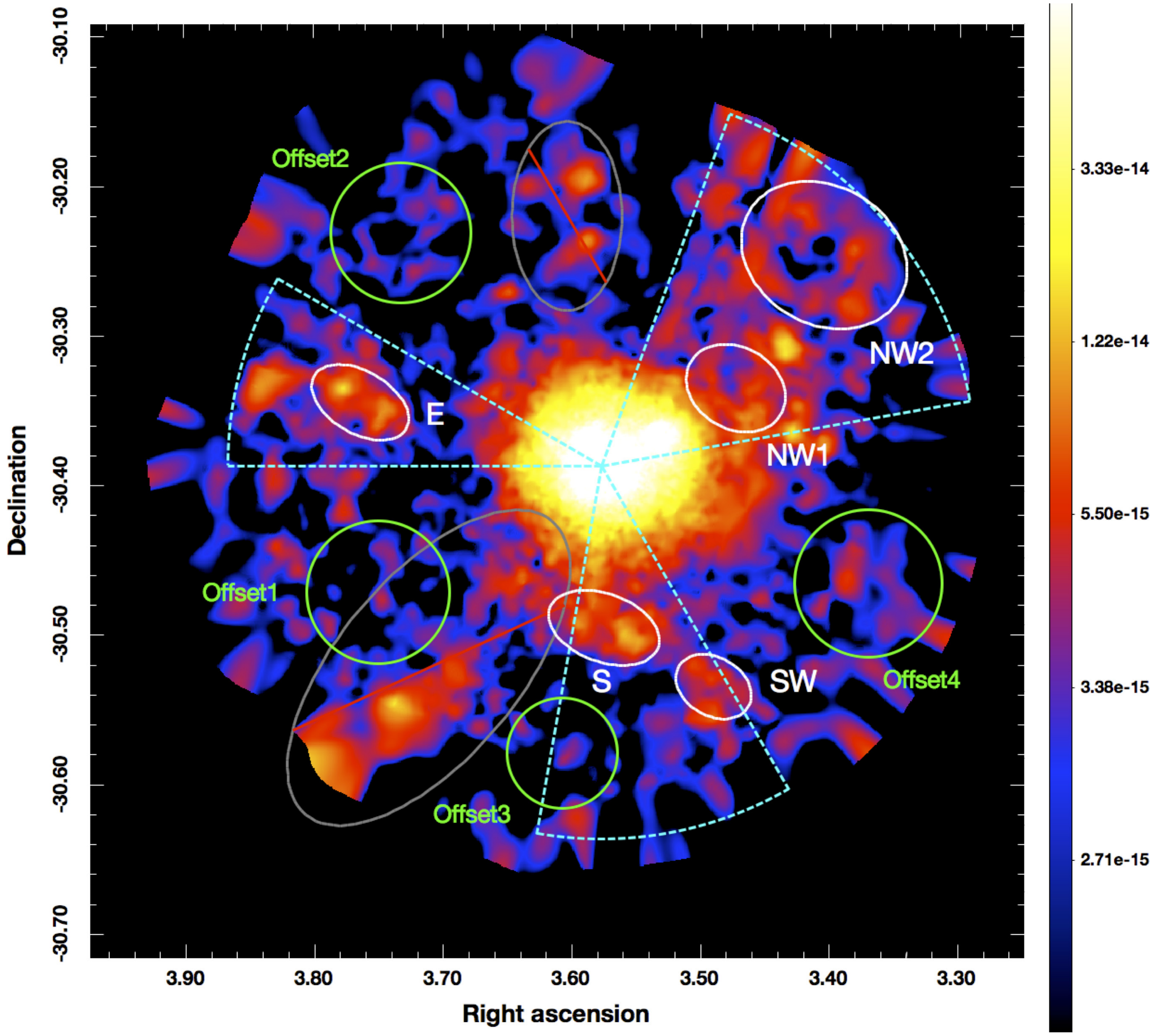
Code availability. The PROFFIT code for X-ray surface brightness analysis is available at http://www.isdc.unige.ch/~deckert/newsite/Proffit.html. The THELI data reduction scheme for CFHT and ESO/WFI data can be downloaded at https://www.astro.uni-bonn.de/theli/. The gravitational lensing code LENSTOOL can be found at http://projets.lam.fr/projects/lenstool/wiki. The KSBf90 code used for weak lensing is available at http://www.roe.ac.uk/~heymans/KSBf90/Home.html.

29. Snowden, S. L., Mushotzky, R. F., Kuntz, K. D. & Davis, D. S. A catalog of galaxy clusters observed by XMM-Newton. *Astron. Astrophys.* **478,** 615–658 (2008).
30. Moretti, A., Campana, S., Lazzati, D. & Tagliaferri, G. The resolved fraction of the cosmic X-ray background. *Astrophys. J.* **588,** 696–703 (2003).
31. Eckert, D., Molendi, S. & Paltani, S. The cool-core bias in X-ray galaxy cluster samples. I. Method and application to HIFLUGCS. *Astron. Astrophys.* **526,** A79 (2011).
32. Eckert, D. *et al.* The stripping of a galaxy group diving into the massive cluster A2142. *Astron. Astrophys.* **570,** A119 (2014).
33. Smith, R. K., Brickhouse, N. S., Liedahl, D. A. & Raymond, J. C. Collisional plasma models with APEC/APED: emission-line diagnostics of hydrogen-like and helium-like ions. *Astrophys. J.* **556,** L91–L95 (2001).
34. Leccardi, A. & Molendi, S. Radial metallicity profiles for a large sample of galaxy clusters observed with XMM-Newton. *Astron. Astrophys.* **487,** 461–466 (2008).

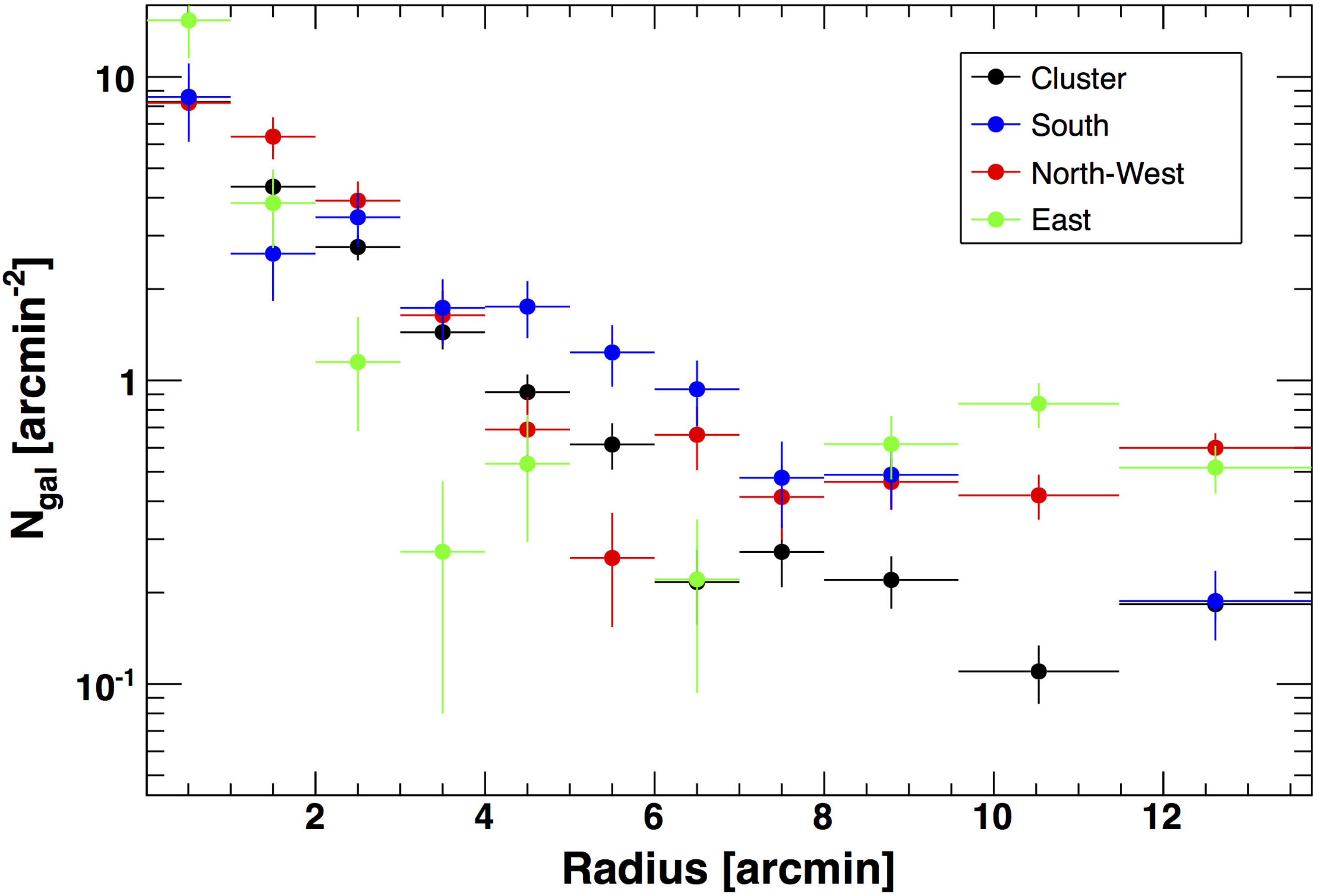
35. Kalberla, P. M. W. *et al.* The Leiden/Argentine/Bonn (LAB) Survey of Galactic HI. Final data release of the combined LDS and IAR surveys with improved stray-radiation corrections. *Astron. Astrophys.* **440,** 775–782 (2005).
36. De Luca, A. & Molendi, S. The 2-8 keV cosmic X-ray background spectrum as observed with XMM-Newton. *Astron. Astrophys.* **419,** 837–848 (2004).
37. McCammon, D. *et al.* A high spectral resolution observation of the soft X-ray diffuse background with thermal detectors. *Astrophys. J.* **576,** 188–203 (2002).
38. Erben, T. *et al.* GaBoDS: The Garching-Bonn Deep Survey. IV. Methods for the image reduction of multi-chip cameras demonstrated on data from the ESO Wide-Field Imager. *Astron. Nachr.* **326,** 432–464 (2005).
39. Schirmer, M. THELI: convenient reduction of optical, near-infrared, and mid-infrared imaging data. *Astrophys. J. Suppl. Ser.* **209,** 21 (2013).
40. Erben, T. *et al.* CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey — imaging data and catalogue products. *Mon. Not. R. Astron. Soc.* **433,** 2545–2563 (2013).
41. Dietrich, J. P. *et al.* BLOX: the Bonn lensing, optical, and X-ray selected galaxy clusters. I. Cluster catalog construction. *Astron. Astrophys.* **470,** 821–834 (2007).
42. Bertin, E. & Arnouts, S. SExtractor: software for source extraction. *Astron. Astrophys. Suppl.* **117,** 393–404 (1996).
43. Jauzac, M. *et al.* A weak lensing mass reconstruction of the large-scale filament feeding the massive galaxy cluster MACS J0717.5+3745. *Mon. Not. R. Astron. Soc.* **426,** 3369–3384 (2012).
44. Leauthaud, A. *et al.* Weak gravitational lensing with COSMOS: galaxy selection and shape measurements. *Astrophys. J. Suppl. Ser.* **172,** 219–238 (2007).
45. Casertano, S. *et al.* WFPC2 observations of the Hubble Deep Field South. *Astron. J.* **120,** 2747–2824 (2000).
46. Rhodes, J., Refregier, A. & Groth, E. J. Weak lensing measurements: a revisited method and application to Hubble Space Telescope images. *Astrophys. J.* **536,** 79–100 (2000).
47. Harvey, D., Massey, R., Kitching, T., Taylor, A. & Tittley, E. The nongravitational interactions of dark matter in colliding galaxy clusters. *Science* **347,** 1462–1465 (2015).
48. Kaiser, N., Squires, G. & Broadhurst, T. A method for weak lensing observations. *Astrophys. J.* **449,** 460–475 (1995).
49. Shan, H. *et al.* Weak lensing measurement of galaxy clusters in the CFHTLS-Wide Survey. *Astrophys. J.* **748,** 56 (2012).
50. Luppino, G. A. & Kaiser, N. Detection of weak lensing by a cluster of galaxies at $z=0.83$. *Astrophys. J.* **475,** 20–28 (1997).
51. Hoekstra, H., Franx, M. & Kuijken, K. Hubble Space Telescope weak-lensing study of the $z=0.83$ cluster MS 1054-03. *Astrophys. J.* **532,** 88–108 (2000).
52. Jullo, E. *et al.* A Bayesian approach to strong lensing modelling of galaxy clusters. *New J. Phys.* **9,** 447–478 (2007).
53. Jauzac, M. *et al.* Hubble Frontier Fields: the geometry and dynamics of the massive galaxy cluster merger MACSJ0416.1-2403. *Mon. Not. R. Astron. Soc.* **446,** 4132–4147 (2015).
54. Jullo, E., Pires, S., Jauzac, M. & Kneib, J.-P. Weak lensing galaxy cluster field reconstruction. *Mon. Not. R. Astron. Soc.* **437,** 3969–3979 (2014).



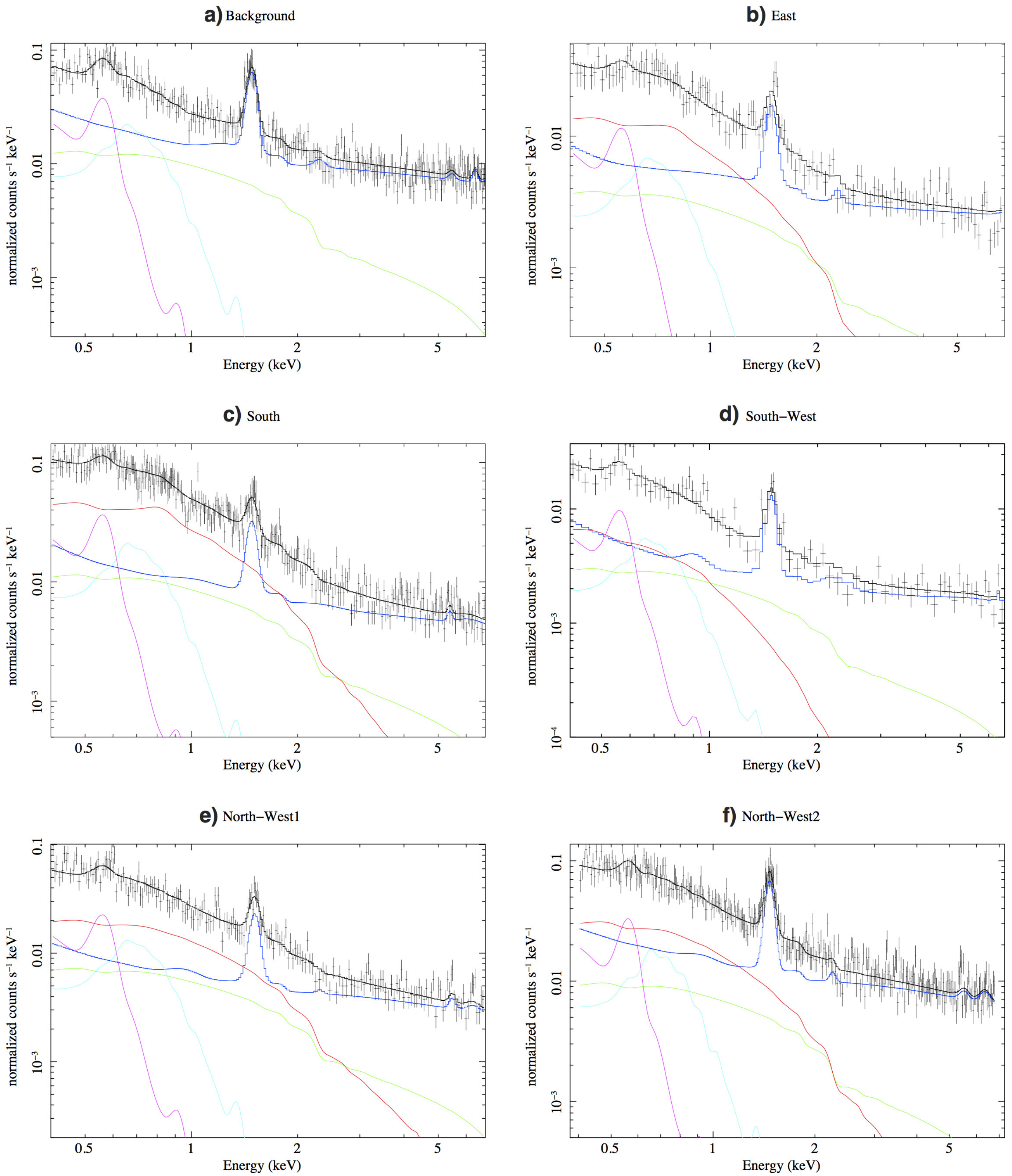
Extended Data Figure 1 | Radial X-ray emissivity profiles in the filaments and in the cluster. Shown are XMM-Newton/EPIC surface-brightness profiles (S_X); black, obtained by masking the filaments; colours, surface brightness in the sectors NW (northwest, position angle 10°–70°), E (east, 150°–180°) and S (south, 260°–300°). Uncertainties (error bars) are given at the 1σ level.



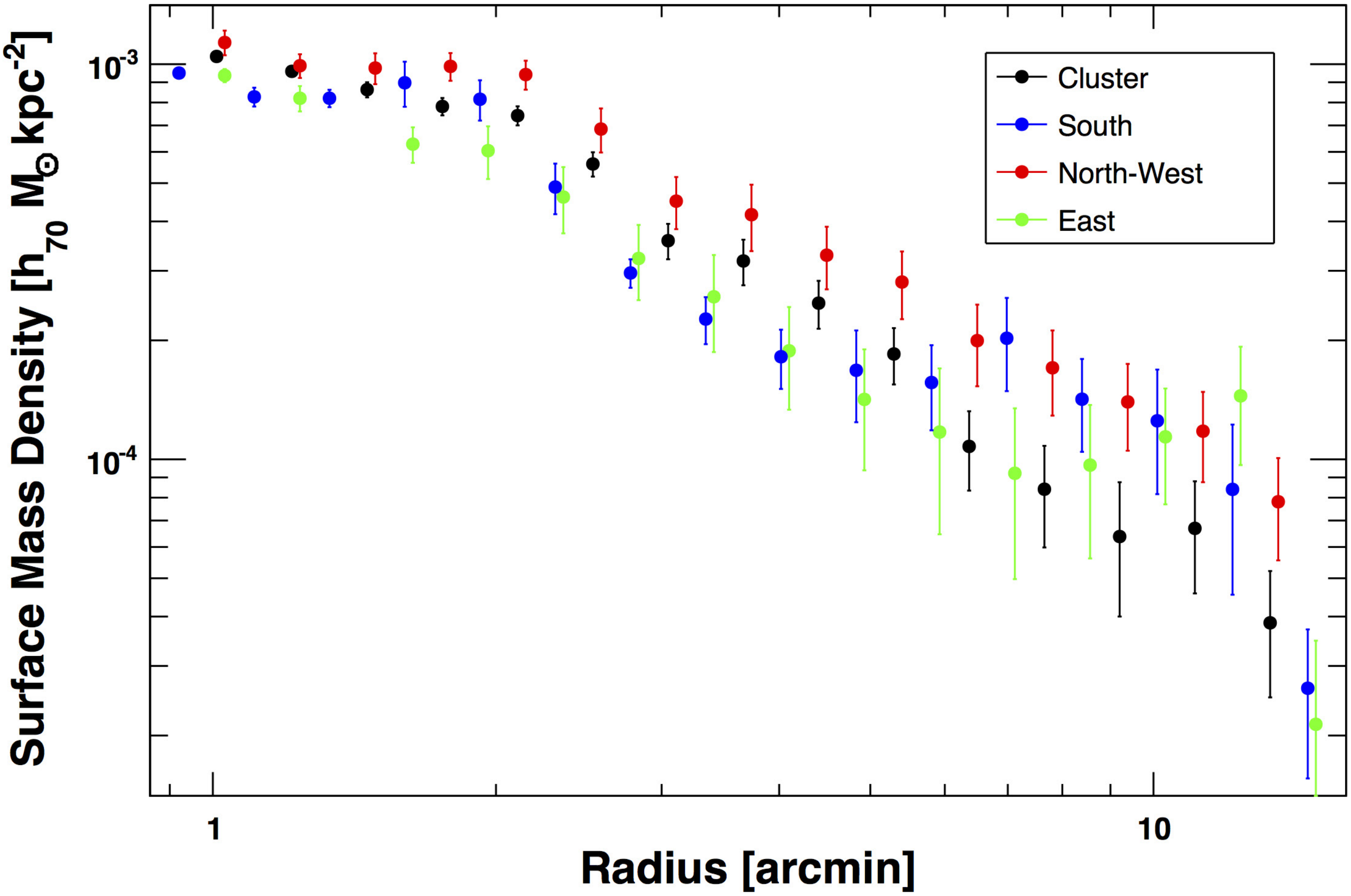
Extended Data Figure 2 | Regions used for the analysis of the thermodynamic properties of the filaments. The 0.5–2 keV surface brightness level is colour coded (bar at right; units are erg s^{-1} cm^{-2} $arcmin^{-2}$); right ascension and declination are in degrees. Spectra were extracted from the regions indicated as E, S, SW, NW1 and NW2 by the white ellipses. The green circles show the regions labelled as Offset1–4 used to estimate the local background components (see Extended Data Table 1). The dashed cyan sectors show the regions used to extract the radial profiles along the filaments for Extended Data Figs 1, 3 and 5. The grey ellipses show background/foreground structures masked during the analysis (see text).



Extended Data Figure 3 | Radial galaxy density profiles in the filaments and in the cluster. Galaxy density profiles (N_{gal}) using spectroscopically confirmed cluster members in sectors encompassing the filaments (same as Extended Data Fig. 1) are compared to the galaxy density of the cluster obtained by masking the filaments (black). Uncertainties (error bars) are given at the 1σ level.



Extended Data Figure 4 | X-ray spectra of the filaments. a–f, XMM-Newton/EPIC-pn spectra for the regions shown in Extended Data Fig. 2. The background region (**a**) refers to Offset1. The fitting procedure was performed jointly on all EPIC instruments; however, only the pn spectra are shown here for clarity. The coloured lines show fitted contributions from the source (red), the NXB (blue), the CXB (green), the Galactic halo (cyan), and the local hot bubble (magenta).



Extended Data Figure 5 | Radial mass profiles in the filaments and in the cluster. Shown are surface mass density profiles obtained from combined strong and weak lensing. The black curve shows the cluster average, compared to the profiles obtained in the direction of the filaments (same as Extended Data Fig. 1).

Extended Data Table 1 | Properties of the X-ray background in the Abell 2744 region

Region	CXB	Halo kT	Halo Norm	LB Norm
Offset 1	$(6.26 \pm 0.56) \times 10^{-7}$	0.297 ± 0.024	$(4.45 \pm 0.60) \times 10^{-7}$	$(1.89 \pm 0.25) \times 10^{-6}$
Offset 2	$(7.03 \pm 0.71) \times 10^{-7}$	0.368 ± 0.095	$(2.31 \pm 0.91) \times 10^{-7}$	$(2.36 \pm 0.36) \times 10^{-6}$
Offset 3	$(6.92 \pm 0.78) \times 10^{-7}$	0.311 ± 0.034	$(5.05 \pm 0.88) \times 10^{-7}$	$(2.14 \pm 0.36) \times 10^{-6}$
Offset 4	$(7.65 \pm 0.71) \times 10^{-7}$	0.283 ± 0.036	$(3.52 \pm 0.82) \times 10^{-7}$	$(2.40 \pm 0.28) \times 10^{-6}$

Comparison of X-ray background parameters per square arcminute obtained in regions Offset 1, 2, 3 and 4 (see Extended Data Fig. 1). CXB, cosmic X-ray background, in photons keV^{-1} cm^{-2} s^{-1} at 1 keV; Halo kT, in keV; Halo Norm, halo normalization, LB Norm, local bubble normalization, both as $\int n_e n_H dV \times 10^{-14}/(4\pi d_A^2(1+z)^2)$, where d_A indicates the angular diameter distance at redshift z.

Extended Data Table 2 | Mass of the filaments

Region	M_{HLT} $[h_{70}^{-1} M_{\odot}]$	S/N	M_{KSB} $[h_{70}^{-1} M_{\odot}]$	S/N
E	$(7.9 \pm 2.8) \times 10^{13}$	3.1	$(4.4 \pm 3.1) \times 10^{13}$	2.1
S	$(9.5 \pm 2.4) \times 10^{13}$	6.8	$(4.0 \pm 2.4) \times 10^{13}$	2.3
SW	$(4.8 \pm 1.7) \times 10^{13}$	3.1	$(2.2 \pm 1.6) \times 10^{13}$	2.8
NW1	$(9.5 \pm 2.7) \times 10^{13}$	5.2	$(6.9 \pm 3.0) \times 10^{13}$	2.2
NW2	$(1.2 \pm 0.3) \times 10^{14}$	3.3	$(2.2 \pm 1.0) \times 10^{14}$	2.6

Comparison of weak-lensing masses for the filaments for the two methods used here: the grid-based multi-scale approach (hybrid LENSTOOL, HLT, giving M_{HLT}) and the direct inversion method (KSB, giving M_{KSB}).

LETTER

doi:10.1038/nature15751

Relativistic baryonic jets from an ultraluminous supersoft X-ray source

Ji-Feng Liu[1,2], Yu Bai[1], Song Wang[1], Stephen Justham[1,2], You-Jun Lu[1,2], Wei-Min Gu[3], Qing-Zhong Liu[4], Rosanne Di Stefano[5], Jin-Cheng Guo[1], Antonio Cabrera-Lavers[6,7], Pedro Álvarez[6,7], Yi Cao[8] & Shri Kulkarni[8]

The formation of relativistic jets by an accreting compact object is one of the fundamental mysteries of astrophysics. Although the theory is poorly understood, observations of relativistic jets from systems known as microquasars (compact binary stars)[1,2] have led to a well established phenomenology[3,4]. Relativistic jets are not expected to be produced by sources with soft or supersoft X-ray spectra, although two such systems are known to produce relatively low-velocity bipolar outflows[5,6]. Here we report the optical spectra of an ultraluminous supersoft X-ray source (ULS[7,8]) in the nearby galaxy M81 (M81 ULS-1; refs 9, 10). Unexpectedly, the spectra show blueshifted, broad Hα emission lines, characteristic of baryonic jets with relativistic speeds. These time-variable emission lines have projected velocities of about 17 per cent of the speed of light, and seem to be similar to those from the prototype microquasar SS 433 (refs 11, 12). Such relativistic jets are not expected to be launched from white dwarfs[13], and an origin from a black hole or a neutron star is hard to reconcile with the persistence of M81 ULS-1's soft X-rays[10]. Thus the unexpected presence of relativistic jets in a ULS challenges canonical theories of jet formation[3,4], but might be explained by a long-speculated, supercritically accreting black hole with optically thick outflows[14–20].

Initial spectroscopic observations[21] of M81 ULS-1, made at the W. M. Keck Observatory in 2010, found broad Balmer hydrogen emission lines (as wide as 400 km s^{-1}) on top of a power-law-like blue continuum. A very broad emission line (as wide as 30 Å, corresponding to 2,000 km s^{-1}) was detected at around 5,532 Å and 5,543 Å in both observations, but was not identified with any known spectral lines. We followed up with new spectra obtained at the Gran Telescopio Canarias in 2015, which again showed the Balmer emission lines and the blue continuum; however, the previously unidentified broad emission line was now at a notably changed wavelength of 5,648 Å (Fig. 1). This change in observer-frame wavelength immediately suggests that the previously unidentified emission line is a blueshifted Hα emission line emitted by an approaching baryonic relativistic jet, at projected velocities of 17% of the speed of light ($-0.17c$). Subsequent spectra reveal ongoing changes in the projected velocity of the blueshifted jet, for which we suggest that the best explanation is jet precession, as observed in the prototype microquasar SS 433.

SS 433 has exhibited time-variable blueshifted and redshifted optical emission lines from its precessing jets, the long-term monitoring of which has revealed[11,12] a precession period of 164 days, and an intrinsic jet velocity of $0.26c$. M81 ULS-1 is only the second microquasar

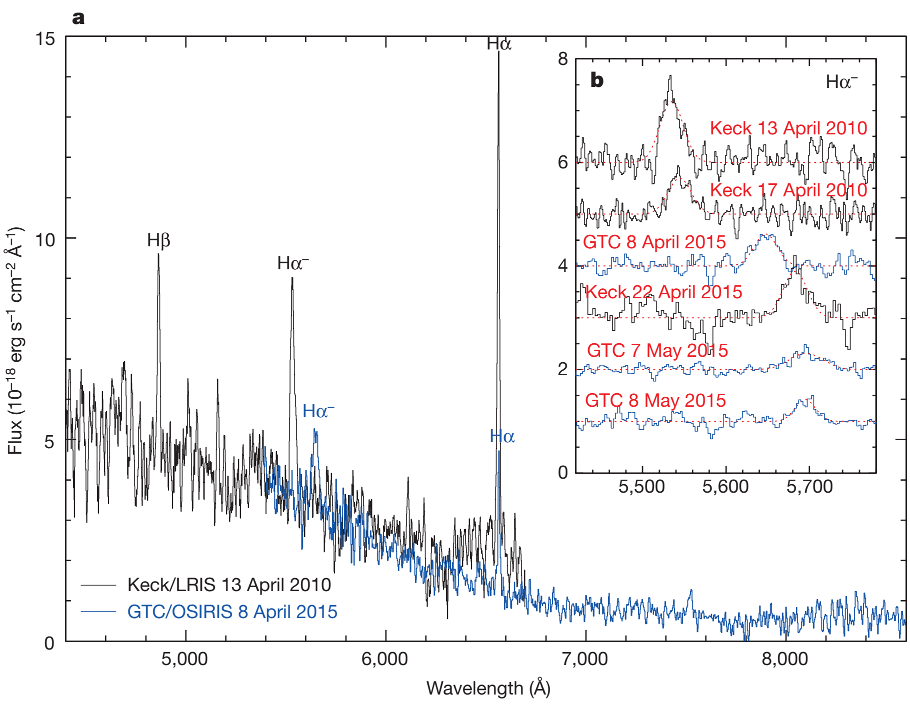


Figure 1 | Spectra obtained from the W. M. Keck Observatory and the Gran Telescopio Canarias (GTC) for the optical counterpart of M81 ULS-1. a, The Keck/LRIS (Low Resolution Imaging Spectrometer) spectrum taken on 13 April 2010 (blue channel; shown in black) and the GTC/OSIRIS (Optical System for Imaging and Low/Intermediate-Resolution Integrated Spectroscopy) spectrum taken on 8 April 2015 (shown in blue) for M81 ULS-1. Labelled are the broad Balmer lines (Hα and Hβ), and the very broad blueshifted $H\alpha^-$ lines at 5,530 Å (Keck/LRIS) and 5,648 Å (GTC/OSIRIS). The power-law-like continuum and the broad Balmer lines are characteristic of an accretion disk around a compact object, confirming the physical association between the X-ray source and its optical counterpart. **b**, The blueshifted $H\alpha^-$ emission lines from six Keck and GTC observations, with time-variable, observer-frame central wavelengths. The intensities also change with time in proportion to the intensities of the stationary Hα emission line from the accretion disk, suggesting a link between the accretion and the jet. See Methods for details. Both the spectra and the fits are normalized by the underlying continuum, and are shifted vertically for clarity.

[1]Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing 100012, China. [2]College of Astronomy and Space Sciences, University of Chinese Academy of Sciences, 19A Yuquan Road, Beijing 100049, China. [3]Department of Astronomy, Xiamen University, Xiamen, Fujian Province 361005, China. [4]Key Laboratory of Dark Matter and Space Astronomy, Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing, Jiangsu Province 210008, China. [5]Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA. [6]Instituto de Astrofísica de Canarias, c/Vía Láctea s/n, E-38200 La Laguna, Tenerife, Spain. [7]Departamento de Astrofísica, Universidad de la Laguna, E-38205 La Laguna, Tenerife, Spain. [8]Department of Astronomy, Caltech, Pasadena, California 91125, USA.

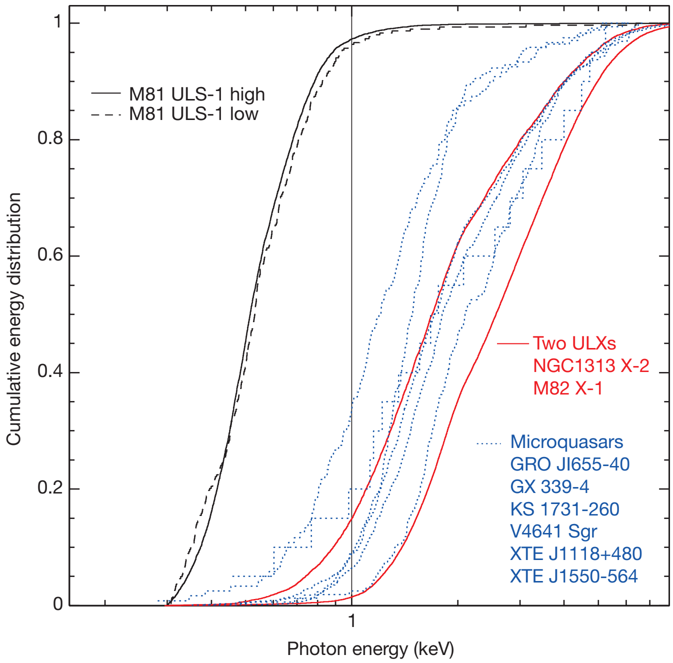


Figure 2 | Cumulative distributions of photon energies for M81 ULS-1. Shown are the energies in the low-flux and high-flux X-ray states. For comparison, we have also plotted the photon energies from two typical ultraluminous X-ray sources (ULXs) in nearby galaxies, and six Galactic microquasars observed with the Advanced CCD Imaging Spectrometer (ACIS) aboard the Chandra X-Ray Observatory. Most photons concentrate at energies where the cumulative distribution rises fastest. For M81 ULS-1, more than 95% of the photons have energies below 1 keV in both the low-flux state (black dashed line) and the high-flux state (black solid line). For both of the ULXs, fewer than 15% of the photons have energies below 1 keV. For the Galactic microquasars, only a few per cent to less than 35 per cent of the photons have energies below 1 keV. Because the response matrix is not well calibrated below 0.3 keV for Chandra/ACIS, only photons with energies greater than 0.3 keV are shown here.

to be identified through measuring directly the blueshifting of Hα lines emitted by its baryonic jets. Other known microquasars[1,2] have mostly been identified through direct imaging of their radio jets, or by interpreting strong non-thermal radio emission as arising from their relativistic jets with velocities above 0.1c. M81 ULS-1 has not previously been detected by radio surveys, but this is not surprising given the great distance to the galaxy M81. Were SS 433 placed in M81, its radio flux at Earth would be about 1 μJy—below the detection sensitivity of current radio facilities, but achievable in the future with the Square Kilometer Array[22].

From its X-ray properties[10], M81 ULS-1 seems to be a truly unique jet source, different to all other known microquasars[1,2]. Since the launch of the Chandra X-Ray Observatory, all observations of M81 have detected ULS-1, which exhibits high-flux and low-flux states with count rates ranging from 1 to 70 photons per kilosecond. When in high-flux states, M81 ULS-1 clearly exhibits supersoft spectra with blackbody temperatures of 65–100 eV and bolometric luminosities greater than 10^{39} erg s^{-1}. Somewhat surprisingly, the low-flux state of M81 ULS-1 appears to be as supersoft as the high-flux state, with more than 95% of the photons having energies below 1 keV (Fig. 2). In contrast, all other known microquasars[1,2] are low-mass or high-mass X-ray binaries, each shown or thought to contain a neutron star or black hole, emitting abundant hard photons with energies above 1 keV. Observations by the Chandra X-Ray Observatory show that only a few per cent to 35 per cent of the photons from these microquasars have energies below 1 keV (Fig. 2).

Luminous supersoft sources[23] have supersoft X-ray spectra, and those with luminosities below the Eddington limit for a solar-mass object are conventionally interpreted as white dwarfs accreting at a rate of about $10^{-7}\,M_\odot\,\mathrm{yr}^{-1}$ to $10^{-6}\,M_\odot\,\mathrm{yr}^{-1}$ (where $M_\odot$ is the mass of the Sun), where hydrogen fusion within the accreted material proceeds steadily[24,25]. But for M81 ULS-1, the presence of relativistic jets suggests otherwise: such jets are simply not expected for typical white dwarfs[13]. Indeed, although bipolar outflows with low velocities of a few thousand kilometres per second are possible and have been observed in supersoft sources such as RX J0513.9-6951 (ref. 5) and RX J0019.8+2156 (ref. 6), no relativistic jets have ever, to our knowledge, been observed from supersoft sources other than M81 ULS-1. These considerations suggest that the accreting object in M81 ULS-1 is not a white dwarf, adding strong evidence to the idea[26,27] that supersoft sources, especially the ultraluminous ones, do not necessarily contain accreting white dwarfs.

If, instead, the central engine of M81 ULS-1 is a neutron star or a black hole—as is the case for all other known microquasars—established phenomenology[3,4] would predict steady jets to be generated when X-ray emissions are in the low-hard state, with episodic jets generated when emissions are in the very high state, or during the transitions between soft and hard states. In the case of M81 ULS-1, the blueshifted Hα emission lines emitted from the relativistic jets were present in all six optical spectroscopic observations in 2010 and 2015. Standard presumptions would therefore be that M81 ULS-1 is in the low-hard or very high states for a substantial fraction of the time, during which abundant hard photons (with energies above 1 keV) would be expected, as for other microquasars (Fig. 2). However, X-ray emissions from ULS-1 have been supersoft in all 19 Chandra observations, regardless of whether ULS-1 is displaying a low-flux or a high-flux state—suggesting that its relativistic jets are not generated in the canonical ways. In fact, the persistently supersoft appearance of ULS-1 would not be expected in any spectral states in the standard accretion scenarios[28,29] for neutron-star or black-hole X-ray binaries, which are known to be accreting below the critical (that is, Eddington) rate.

This unusual combination of relativistic jets and persistently supersoft X-ray spectra is completely unexpected, posing a challenge to the conventional understanding of jet formation[3,4]. One possible identity for M81 ULS-1 is a long-speculated[14,15], supercritically accreting black hole with optically thick outflows. Recent magnetohydrodynamic simulations of such systems, although still under development and the subject of heated debate[16,17], can generate super-Eddington luminosities, and necessarily[30] generate disk winds and funnels along the rotation axis, from which radiation pressure will drive baryon-loaded relativistic jets with velocities of up to 0.3c, regardless of the black-hole spin[17]. Observations of M81 ULS-1 qualitatively match the predictions of high luminosities and baryon-loaded relativistic jets, and its supersoft X-ray spectra might be expected from optically thick outflows under suitable conditions of outflow geometry, wind velocities and outflow mass rates[18–20]. Thus, ULS-1 might be a manifestation of recent predictions of supercritical accretion onto black holes, and so reveal the nature of extreme accretion in extreme conditions.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 June; accepted 11 September 2015.
Published online 25 November 2015.

1. Mirabel, I. F. & Rodriguez, L. F. Microquasars in our Galaxy. *Nature* **392,** 673–676 (1998).
2. Paredes, J. M. & Marti, J. Microquasars in the Galaxy. *Contrib. Sci.* **2,** 303–314 (2003).
3. Fender, R. P., Belloni, T. M. & Gallo, E. Towards a unified model for black hole X-ray binary jets. *Mon. Not. R. Astron. Soc.* **355,** 1105–1118 (2004).
4. Migliari, A. & Fender, R. P. Jets in neutron star X-ray binaries: a comparison with black holes. *Mon. Not. R. Astron. Soc.* **366,** 79–91 (2006).
5. Southwell, K. A., Livio, M., Charles, P. A., O'Donoghue, D. & Sutherland, W. J. The nature of the supersoft X-ray source RX J0513–69. *Astrophys. J.* **470,** 1065–1074 (1996).
6. Becker, C. M., Remillard, R. A., Rappaport, S. A. & McClintock, J. E. Bipolar jets and orbital dynamics of the supersoft X-ray source RX J0019.8+2156. *Astrophys. J.* **506,** 880–891 (1998).

7. Di Stefano, R. & Kong, A. K. H. Luminous supersoft X-ray sources in external galaxies. *Astrophys. J.* **592,** 884–899 (2003).
8. Swartz, D. A., Ghosh, K. K., Suleimanov, V., Tennant, A. F. & Wu, K. Chandra discovery of luminous supersoft X-ray sources in M81. *Astrophys. J.* **574,** 382–397 (2002).
9. Liu, J.-F. & Di Stefano, R. An ultraluminous supersoft X-ray source in M81: an intermediate-mass black hole? *Astrophys. J.* **674,** L73–L76 (2008).
10. Liu, J.-F. No periodicity revealed for an "eclipsing" ultraluminous supersoft X-ray source in M81. *Astrophys. J.* (Suppl.) **177,** 181–188 (2008).
11. Margon, B. Observations of SS 433. *Annu. Rev. Astron. Astrophys.* **22,** 507–536 1984).
12. Blundell, K. M., Bowler, M. G. & Schmidtobreick, L. Fluctuations and symmetry in the speed and direction of the jets of SS 433 on different timescales. *Astron. Astrophys.* **474,** 903–910 (2007).
13. Livio, M. Astrophysical jets. In *Proc. Astron. Soc. Pacific Conf. on Probing the Physics of Active Galactic Nuclei* (eds Peterson, B. M., Pogge, R. W. & Polidan, R. S.) **224,** 225–248 (Astron. Soc. Pacific, 2001).
14. Paczýnski, B. & Wiita, P. J. Thick accretion disks and supercritical luminosities. *Astron. Astrophys.* **88,** 23–31 (1980).
15. Abramowicz, M. A., Czerny, B., Lasota, J. P. & Szuszkiewicz, E. Slim accretion disks. *Astrophys. J.* **332,** 646–658 (1988).
16. Jiang, Y.-F., Stone, J. M. & Davis, S. W. A global three-dimensional radiation magneto-hydrodynamic simulation of super-Eddington accretion disks. *Astrophys. J.* **796,** 106 (2014).
17. Sadowski, A. & Narayan, R. Powerful radiative jets in super-critical accretion disks around non-spinning black holes. Preprint at http://arxiv.org/abs/1503.00654 (2015).
18. Ohsuga, K., Mori, M., Nakamoto, T. & Mineshige, S. Supercritical accretion flows around black holes: two-dimensional, radiation pressure-dominated disks with photon trapping. *Astrophys. J.* **628,** 368–381 (2005).
19. King, A. & Pounds, K. Black hole winds. *Mon. Not. R. Astron. Soc.* **345,** 657–659 (2003).
20. Shen, R.-F., Duran, R. B., Nakar, E. & Piran, T. The nature of ULX source M101 ULX-1: optically thick outflow from a stellar mass black hole. *Mon. Not. R. Astron. Soc.* **447,** L60–L64 (2015).
21. Bai, Y., Liu, J. & Wang, S. Spectroscopic studies of an ultraluminous supersoft X-ray source in M81. *Astrophys. J.* **802,** L27 (2015).
22. Schilizzi, R. T. *et al.* Preliminary specifications for the Square Kilometre Array. https://www.skatelescope.org/uploaded/5110_100_Memo_Schilizzi.pdf (2007).
23. Kahabka, P. & van den Heuvel, E. P. J. Luminous supersoft X-ray sources. *Annu. Rev. Astron. Astrophys.* **35,** 69–100 (1997).
24. Iben, I. Jr Hot accreting white dwarfs in the quasi-static approximation. *Astrophys. J.* **259,** 244–266 (1982).
25. Nomoto, K. Accreting white dwarf models for type I supernovae. I. Presupernova evolution and triggering mechanisms. *Astrophys. J.* **253,** 798–810 (1982).
26. Di Stefano, R. & Kong, A. The discovery of quasi-soft and supersoft sources in external galaxies. *Astrophys. J.* **609,** 710–727 (2004).
27. Di Stefano, R., Primini, F. A., Liu, J.-F., Kong, A. & Patel, B. Populations of supersoft X-ray sources: novae, tidal disruption, type Ia supernovae, accretion-induced collapse, ionization, and intermediate-mass black holes? *Astron. Nachr.* **331,** 205–211 (2010).
28. Remillard, R. A. & McClintock, J. E. X-ray properties of black-hole binaries. *Annu. Rev. Astron. Astrophys.* **44,** 49–92 (2006).
29. Hasinger, G. & van der Klis, M. Two patterns of correlated X-ray timing and spectral behaviour in low-mass X-ray binaries. *Astron. Astrophys.* **225,** 79–96 (1989).
30. Gu, W.-M. Mechanism of outflows in accretion system: advective cooling cannot balance viscous heating? *Astrophys. J.* **799,** 71 (2015).

Acknowledgements We thank K. Blundell, R. Narayan, Z. Li, T. Wang, F. Yuan, X. Fang, J. Irwin, T. Maccarone and D. Swartz for helpful discussions. We acknowledge support from the Chinese Academy of Sciences (grant XDB09000000), from the 973 Program (grant 2014CB845705), and from the National Science Foundation of China (grants NSFC-11333004/11425313). This work is based partly on observations made with the Gran Telescopio Canarias, installed in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias, on the island of La Palma. Some of the data were obtained at the W. M. Keck Observatory, which is operated through a scientific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. This Observatory was made possible through the financial support of the W. M. Keck Foundation.

Author Contributions J.-F.L. proposed the observations. J.-F.L., Y.B., S.W. and J.-C. G. reduced the optical and X-ray data and carried out the analysis. J.-F.L., S.J., Y.-J.L. and R.D.S. discussed the results and drafted the manuscript. A. C.-L., P.A., Y.C and S.K. helped with the observations. All authors commented on and helped in improving the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-F. Liu (jfliu@nao.cas.cn).

METHODS

GTC/OSIRIS and Keck/LRIS data reduction. Initial optical spectroscopic observations of M81 ULS-1 were carried out with Keck/LRIS on 13 April and 17 April 2010, revealing broad Balmer emission lines as if from an accretion disk[21]. The blueshifted Hα emission line (Hα$^-$) is shown at 5,530 Å in the spectra, with a shift in the line centre of 10 Å ± 2 Å between those two observations (that is, two epochs separated by four nights).

M81 ULS-1 was later observed using GTC/OSIRIS on 8 April, 7 May and 8 May 2015, masked with the 0.6″ slit followed by the R1000R grating, which yields a resolution of about 7 Å. The spectra were reduced in a standard way with IRAF (Image Reduction and Analysis Facility) software (http://iraf.noao.edu). After bias subtraction and flat correction, dispersion correction was carried out on the basis of the line lists given in the OSIRIS manual (http://www.gtc.iac.es/instruments/osiris/). Raw spectra were then extracted with an aperture size of 1″, and a standard star taken at each night was used to make the flux calibration.

On 22 April 2015, another observation of M81 ULS-1 was carried out using Keck/LRIS with the 1.0″ slit. The light was split with a beam dichroic of 6,800 Å to the blue and red sides, followed by using the 300/5,000 and 400/8,500 gratings, which yields a resolution of ∼8 Å. The spectrum was reduced with the IDL (Interactive Data Language) pipeline designed for the W. M. Keck Observatory.

Extended Data Table 1 lists the basic information obtained from the 2010 and 2015 observations. Both Hα and Hα$^-$ emission lines are detected in all the spectra, and their line properties are calculated from Gaussian line profile fitting. Extended Data Table 2 lists the central wavelength, the full width at half-maximum (FWHM) and the equivalent width for each fitted emission line. The observed Hα$^-$ central wavelengths, λ_-, correspond to projected velocities, v_r, from $-0.17c$ to $-0.14c$ in these observations, given by $\lambda_-/\lambda_0 = \sqrt{\frac{1+v_r/c}{1-v_r/c}}$.

Properties of the emission lines. In the case of SS 433, the equivalent widths of Hα emission lines are tightly correlated with the phases of the precession, and those of the Hα$^-$ emission follow a similar trend but with a phase delay[31]. We use the power of the emission lines, calculated from the area of the Gaussian fitting, as representative of the emission intensity, because the observed continuum from M81 ULS-1 varies markedly between observations. Extended Data Fig. 1 shows that the power of the Hα emission lines from the accretion disk is positively correlated with that of the Hα$^-$ emission lines, suggesting a link between the accretion and the jet. The variations in the power of the emission lines are asymmetrical, with smooth rises and steeper declines around 7 May 2015 (Extended Data Fig. 2)—similar to the variations seen in the SS 433 emission lines[31].

The rate at which the projected Hα$^-$ velocity changes seems to be slower during 2015 than it was during 2010. The rate of change in 2015 was roughly 0.8 Å per day, whereas it was 2.6 Å per day in 2010; if the velocity shift is due to precession, this difference may be explained naturally, because the 2015 observations are sampling a different part of the precession cycle. We can estimate a minimum likely precession period by assuming that the turning point of the precession cycle occurred at around the time of the observations of 7 and 8 May 2015 (see Extended Data Figs 2, 3). If so, then, after the wavelength of the emission lines reached the maximum on 8 May (Extended Data Fig. 2), the Hα$^-$ emission probably turned back to the short wavelength with the rate of roughly 0.8 Å per day, indicating that the half-precession period must be longer than 30 days.

There is a 115-Å gap between the observations of 13 April 2010 and 8 April 2015 (Extended Data Fig. 3), and if we assume that the maximum rate at which the wavelength decreases is 2.6 Å per day, then the Hα$^-$ line needs 44 days to move by the required amount. Therefore, the half-precession period is probably longer than 74 days, and lower limit of the precession period is about 148 days. More time-resolved spectra are needed in order to derive an accurate period and to characterize further the apparent precession of the jets.

Searching for the redshifted Hα emission lines. Given the existence of blueshifted Hα$^-$ emission lines from the approaching jets, redshifted Hα emission lines (Hα$^+$) would be expected from receding jets, albeit with much lower intensities (because of Doppler boosting effects). Assuming symmetrical and steady jets, the boosting factors (D) for the lines emitted from the approaching and the receding jets are given by $D_- = \frac{(1-\beta^2)^{\frac{1}{2}}}{1-\beta\cos\theta} > 1$ and $D_+ = \frac{(1-\beta^2)^{\frac{1}{2}}}{1+\beta\cos\theta} < 1$, respectively. The total flux of a blueshifted or redshifted line in the observer frame is boosted by a factor of D^3. The expected central wavelengths of the two lines are given by $\lambda_- = \lambda_0/D_-$, and $\lambda_+ = \lambda_0/D_+$, and the corresponding redshifts are $z_+ = \lambda_+/\lambda_0 - 1 > 0$ and $z_- = \lambda_-/\lambda_0 - 1 < 0$. D and z values for Hα$^-$ in all observations are listed in Extended Data Table 3.

We have searched for the redshifted Hα$^-$ emission lines in all observations. A weak emission line feature was detected at ∼3σ at around 7,524 Å (Fig. 1), roughly symmetrical to Hα$^-$ at 5,648 Å, in one of the GTC exposures during the night of 8 April 2015. If this marginal detection were the redshifted Hα$^-$ line, its boosting factor would be $D_+ = \lambda_0/\lambda_+ = 0.8722$, and the ratio of the received total flux of the blueshifted line to that of the redshifted line should be $D_-^3/D_+^3 = \sim 2.5$, which is roughly consistent with the observed flux ($\sim 4^{+3.5}_{-1.5}$).

However, the observed wavelength is not consistent with the expected Hα$^-$ wavelength given the blueshifted Hα$^-$ at 5,648 Å, that is, $6{,}563\left[\frac{1-\beta\cos\theta}{(1-\beta^2)^{\frac{1}{2}}}\right]$. Assuming the extreme case, $\theta = 0°$, then we have $\beta = 0.1491$. If the receding jet has the same velocity, then the expected central wavelength of Hα$^+$ should be 7,626.4 Å, which is about 104 Å larger than the detected line. If we assume that $\theta = 10°/20°/30°$, then $\beta = 0.152/0.160/0.177$, and the expected central wavelength of Hα$^+$ is 7,632 Å/7,650 Å/7,688 Å, which is about 108 Å/126 Å/164 Å larger than the detected line. The discrepancy becomes larger for larger inclination angles.

This casts doubt on the identification of the 7,524-Å line feature as Hα$^+$, unless the jets are asymmetrical or fast-changing. We may have not detected the redshifted Hα$^+$, but the non-detection is not surprising given the Doppler boosting effects, and other realistic explanations. For example, the receding jets may be blocked by the optically thick outflows if this system is a supercritically accreting black-hole system, as described in the text. No candidate Hα$^+$ emission lines were detected at all in the 7 and 8 May GTC observations, or in the Keck spectrum. Even if the 8 April line were a true Hα$^+$ emission line, this non-detection would not be surprising, given the lower equivalent widths of Hα$^-$ on 7 and 8 May, and the relatively lower sensitivity in the red channel of LRIS.

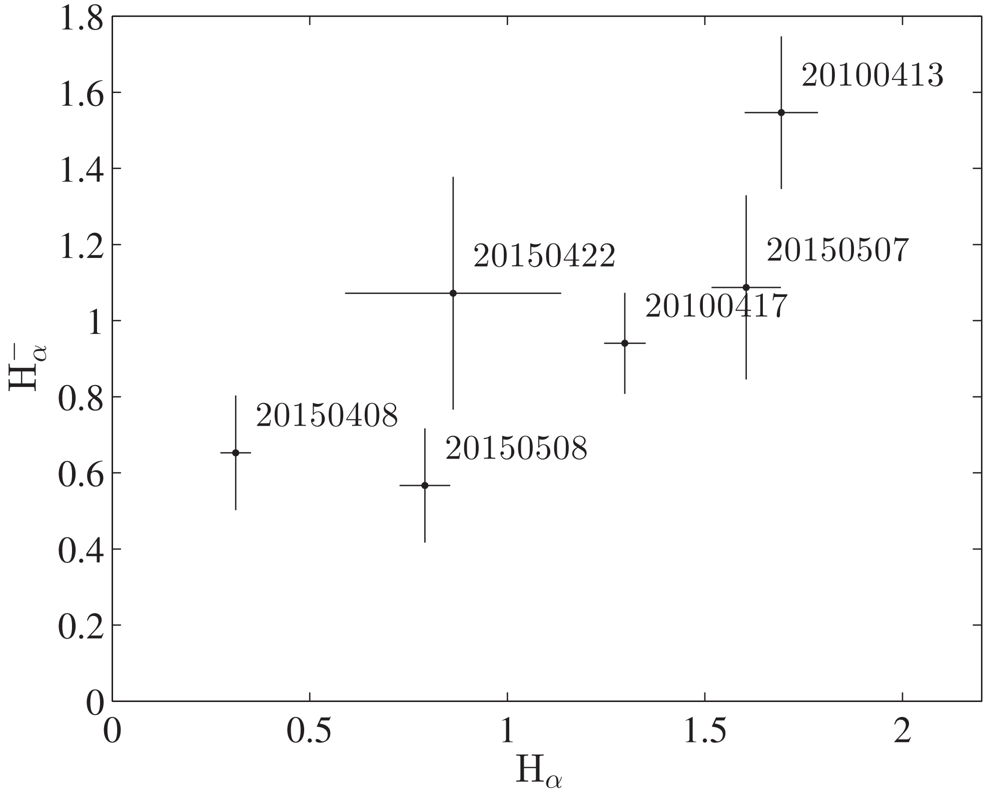
Analysis of Chandra data. There have been 19 Chandra/ACIS observations of the nuclear region of M81, where ULS-1 resides. All of these observations were derived from the Chandra archive and analysed uniformly with CIAO 4.7 software tools (http://cxc.harvard.edu/ciao/). Point sources were detected with WAVDETECT on the individual Chandra images. As listed in Extended Data Table 4, the photon counts were extracted from the source ellipses enclosing 95% of the total photons as reported by WAVDETECT, which was run with scales of 1″, 2″, 4″ and 8″ in the 0.3 to 8.0 keV band.

The spectra in the high-flux states (>10 counts per kilosecond) were fitted by absorbed blackbody models, with the spectral parameters presented in Extended Data Table 4, all of which show that M81 ULS-1 has been persistently supersoft in these observations. In addition, the spectra in the high-flux and low-flux states were added together into combined high- and low-state spectra, and were also fitted in the band 0.3–8.0 keV. Using the fitted absorbed blackbody model, we calculated the 0.3–8.0 keV flux, the 0.3–8.0 keV luminosity and the bolometric luminosity with the distance of 3.63 megaparsecs for M81 (ref. 32).

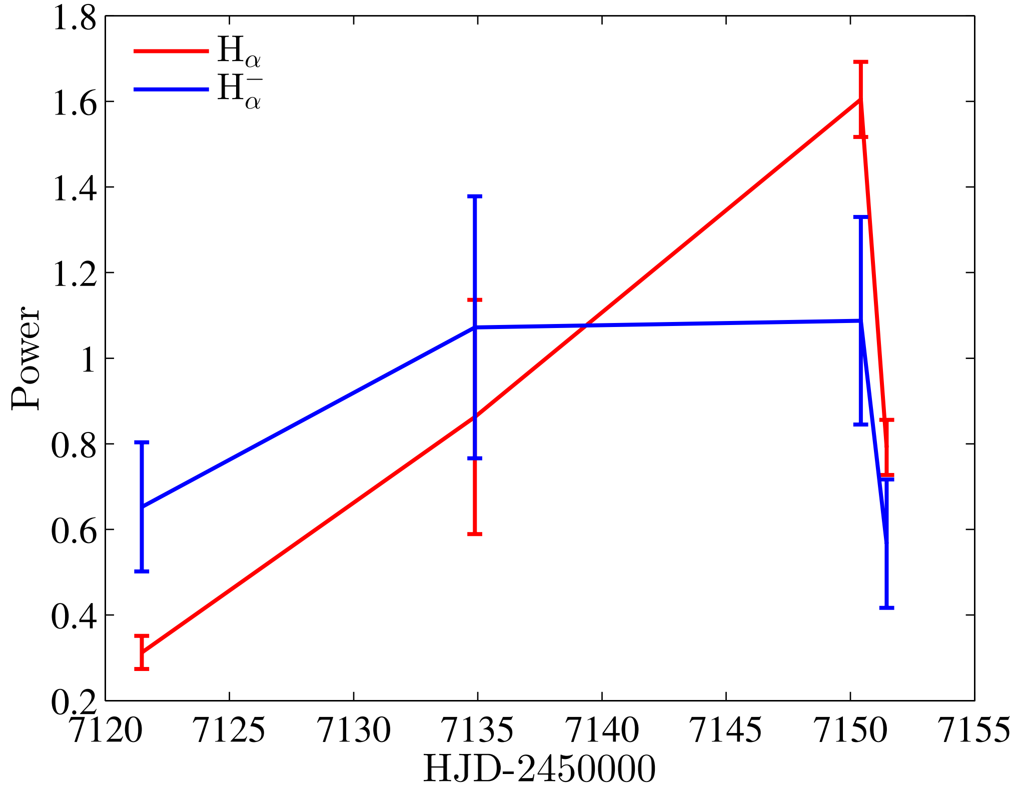
As plotted in Extended Data Fig. 4, M81 ULS-1 displays a soft excess below 0.3 keV as compared to the best-fit model for 0.3–8.0 keV. However, considering that the response matrix for Chandra observations is not well calibrated below 0.3 keV, we refrain from interpreting this soft excess. Nonetheless, it is clear that M81 ULS-1 has very different spectral properties from the other known microquasars. Moreover, these uncertainties in calibration below 0.3 keV might merely make the intrinsic spectral differences between M81 ULS-1 and the other known microquasars even larger (that is, the energy distribution from M81 ULS-1 might be even softer than observed).

Code availability. The optical spectra were reduced with IRAF, available at http://iraf.noao.edu/. All of the emission lines in Extended Data Table 2 were fitted with the curve-fitting toolbox based on Matlab (http://www.mathworks.com/help/curvefit/index.html). The Chandra archive data were analysed with CIAO 4.7, which can be downloaded from http://cxc.harvard.edu/ciao/download/.

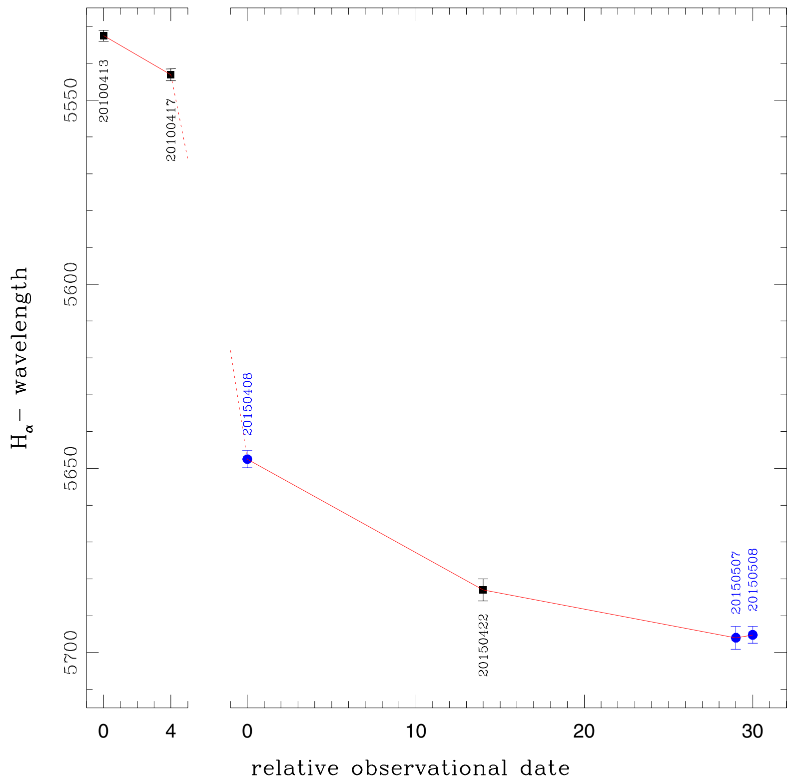
31. Vittone, A., Rusconi, L., Sedmak, G., Mammano, A. & Ciatti, F. Correlations and periodicities of equivalent widths in SS 433. *Astron. Astrophys.* (Suppl.) **53,** 109–117 (1983).
32. Freedman, W. L. *et al.* The Hubble Space Telescope extragalactic distance scale key project. 1. The discovery of Cepheids and a new distance to M81. *Astrophys. J.* **427,** 628–655 (1994).



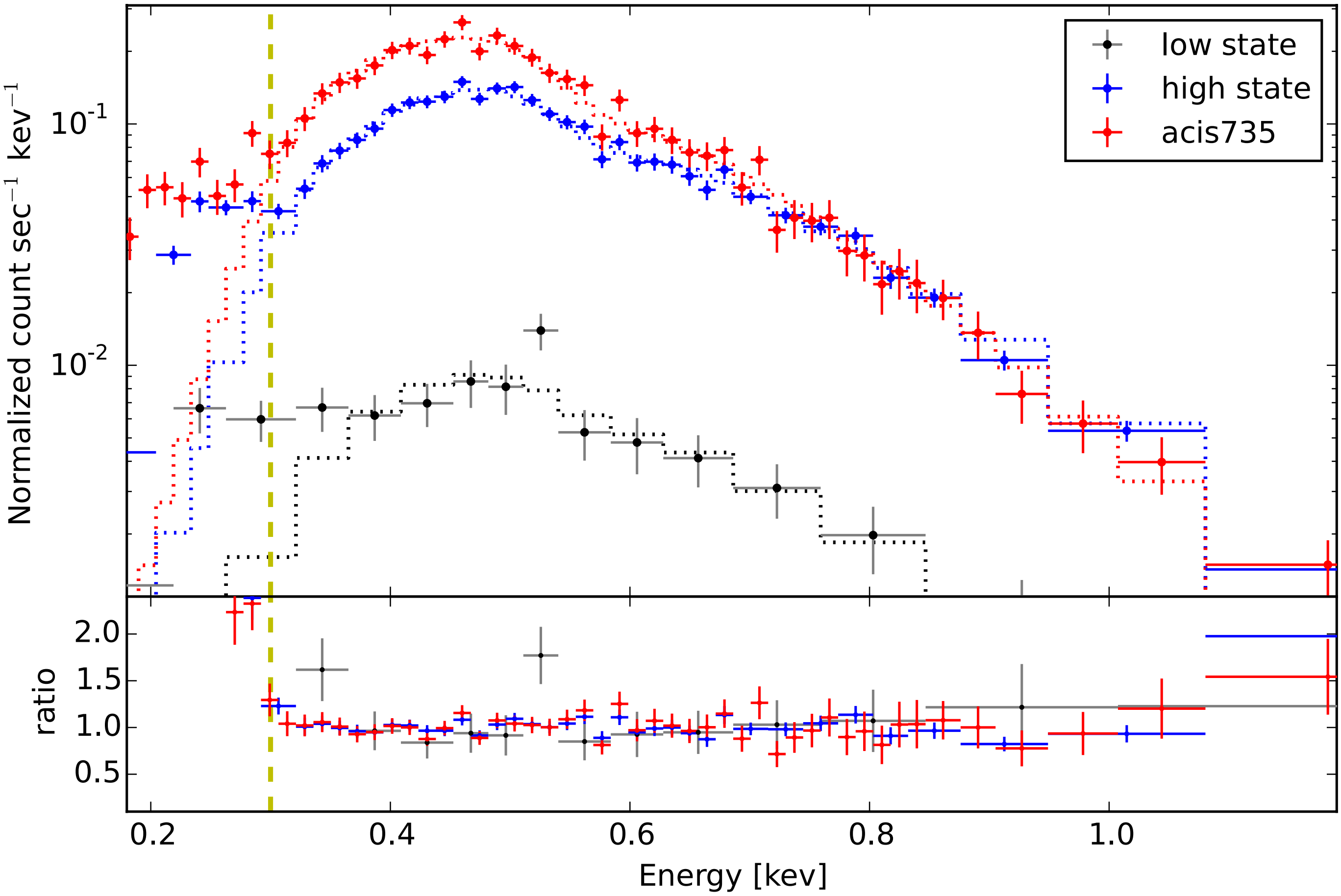
Extended Data Figure 1 | The power of Hα^- emission versus that of Hα emission for M81 ULS-1, in units of 10^{-16} erg s^{-1} cm^{-2}. The error bars denote 68.3% uncertainty. Observations are given as year followed by month followed by day.



Extended Data Figure 2 | Variation in the power of emission lines from M81 ULS-1, in units of 10^{-16} erg s^{-1} cm^{-2}. The error bars denote 68.3% uncertainty. The *x*-axis gives the observation date as a Heliocentric Julian Date (HJD).



Extended Data Figure 3 | The centre of Hα$^-$ emission as a function of the relative observational date. The dates of observations are marked in the figure; the 'relative observational date' refers to the date relative to the first observation of 2010 or 2015. The error bars denote 68.3% uncertainty.



Extended Data Figure 4 | The M81 ULS-1 spectra and model fitting. The spectra from Chandra observation ID 735, the combined high-state observations, and the combined low-state observations are shown with red, blue, and black crosses respectively. The corresponding blackbody models in the energy range 0.3–8.0 keV are shown with red, blue, and black dotted lines. The yellow dashed line indicates photon energy of 0.3 keV. The error bars denote 68.3% uncertainty.

Extended Data Table 1 | Observations of M81 ULS-1

Date	Telescope	Exposure Time (second)	$\Delta\lambda$ (Å)
2010.4.13	Keck	1000×3	5
2010.4.17	Keck	1200×2	5
2015.4.08	GTC	1800×3	7
2015.4.22	Keck	2800×2	8
2015.5.07	GTC	1800×3	7
2015.5.08	GTC	1800×4	7

$\Delta\lambda$ indicates spectral resolution.

Extended Data Table 2 | Properties of Hα^+/Hα^- and Hα for M81 ULS-1

HJD-2450000	H_α^{-+}				H_α		
	Center	FWHM	E.W.	Power	Center	FWHM	Power
5299.83	5532.3 ± 1.5 ($-0.17c$)	33 ± 4	41 ± 5	1.55 ± 0.20	6562.9 ± 0.2	10.8 ± 0.5	1.69 ± 0.09
5303.77	5543.1 ± 1.6 ($-0.17c$)	33 ± 4	24 ± 3	0.94 ± 0.13	6562.8 ± 0.2	9.4 ± 0.3	1.30 ± 0.05
7121.47	5647.5 ± 2.3 ($-0.15c$)	32 ± 6	21 ± 5	0.65 ± 0.15	6564.9 ± 0.3	6.7 ± 0.6	0.31 ± 0.04
7134.88	5683.0 ± 3.0 ($-0.14c$)	34 ± 7	33 ± 9	1.07 ± 0.31	6564.4 ± 0.9	8.9 ± 2.1	0.86 ± 0.27
7150.42	5696.0 ± 3.1 ($-0.14c$)	46 ± 8	16 ± 4	1.08 ± 0.24	6564.1 ± 0.1	7.6 ± 0.3	1.60 ± 0.09
7151.46	5695.2 ± 2.3 ($-0.14c$)	29 ± 6	13 ± 3	0.57 ± 0.15	6564.4 ± 0.2	7.4 ± 0.5	0.79 ± 0.06
7121.47	7522.1 ± 2.7 ($+0.14c$)	20 ± 6	27 ± 7	0.16 ± 0.05			

The centre, FWHM and equivalent width are in units of angströms. The numbers in parentheses are velocities, in units of the speed of light (c). The power is in units of 10^{-16} erg s^{-1} cm^{-2}. All of the error bars denote 68.3% uncertainty. The bottom row shows Hα^+ emission; the other rows show Hα^- emission.

Extended Data Table 3 | Doppler boost factors for each observation of M81 ULS-1

Date	$\lambda(\mathrm{H}_\alpha^-$ or $\mathrm{H}_\alpha^+?)$	D	z
20100413	5532.5	1.1862	-0.1570
20100417	5543.1	1.1840	-0.1554
20150408	5647.5	1.1621	-0.1395
20150422	5683.0	1.1548	-0.1341
20150507	5696.0	1.1522	-0.1321
20150508	5695.2	1.1523	-0.1322
20150408	7524.0	0.8722	0.1465

The second column gives the wavelength of the blueshifted/redshifted Hα emission. *D* is the Doppler boost factor; *z* is the redshift.

Extended Data Table 4 | Chandra observations of M81 ULS-1

ObsID	Obs Date	ExpT (ks)	$C_{\rm NET}$	$C_{\rm Soft}$	Count Rate (count ks^{-1})	$kT_{\rm bb}$ (eV)	$n_{\rm H}$ (10^{20} cm^{-2})	Flux (ergs s^{-1} cm^{-1})	L_X (10^{38} ergs s^{-1})	$L_{\rm bol}$ (10^{38} ergs s^{-1})	χ^2_ν/dof	State
acis390	2000 Mar 21	2.4	140	25	60.81± 5.74	145±40.9	12.5±20.0	1.42e-13	2.3	6.7	1.390/4	high
acis735	2000 May 07	50.7	3679	2141	67.16± 1.19	78±1.6	9.5±1.0	1.99e-13	3.1	23.1	1.086/41	high
acis5935	2005 May 26	11.1	11	2	1.06 ± 0.43							low
acis5936	2005 May 28	11.6	11	2	0.82 ± 0.40							low
acis5937	2005 Jun 01	12.2	22	6	1.63 ± 0.46							low
acis5938	2005 Jun 03	12.0	485	185	37.69± 1.87	81±6.2	12.0±5.8	1.89e-13	3.0	25.6	2.041/17	high
acis5939	2005 Jun 06	12.0	364	181	27.53± 1.61	76±5.6	8.3±4.2	1.62e-13	2.6	17.7	0.857/13	high
acis5940	2005 Jun 09	12.1	70	20	4.90 ± 0.73							low
acis5941	2005 Jun 11	12.0	429	187	32.16± 1.74	91±5.8	6.5±4.0	1.74e-13	2.8	10.6	1.004/17	high
acis5942	2005 Jun 15	12.1	405	206	30.30± 1.68	70±5.4	14.1±5.1	1.68e-13	2.7	42.5	0.724/14	high
acis5943	2005 Jun 18	12.2	525	187	41.30± 1.94	91±5.4	8.8±3.8	2.10e-13	3.3	16.0	1.096/21	high
acis5944	2005 Jun 21	12.0	356	85	28.71± 1.64	96±9.8	20.6±8.6	1.16e-13	1.8	20.3	1.128/15	high
acis5945	2005 Jun 24	11.7	415	220	32.04± 1.75	65±4.4	19.8±5.3	1.69e-13	2.7	97.4	1.028/14	high
acis5946	2005 Jun 26	12.2	287	167	20.75± 1.40	70±7.1	10.6±5.9	1.29e-13	2.0	22.5	1.160/8	high
acis5947	2005 Jun 29	10.8	40	29	2.81 ± 0.62							low
acis5948	2005 Jul 03	12.2	77	51	4.93 ± 0.73							low
acis5949	2005 Jul 06	12.2	54	33	3.36 ± 0.62							low
acis9805	2007 Dec 21	5.2	44	28	7.55 ± 1.43							low
acis9122	2008 Feb 01	10.0	55	18	5.40 ± 0.87							low
Total high		149.3	7085	3584	37.85± 2.53	84±1.3	8.2±0.1	1.98e-13	3.1	16.9	1.739/31	
Total low		97.4	384	189	3.61± 0.81	82±8.3	5.6±11.3	2.03e-14	0.3	1.4	1.069/12	

Column 1 shows the observation identification number. Column 2 shows the observation date. Column 3 shows the on-time (exposure time, ExpT) without dead-time correction. Column 4 shows the net number of photon counts in the range 0.1 to 8.0 keV ($C_{\rm NET}$). Column 5 shows counts in the supersoft band ($C_{\rm Soft}$), 0.1–0.5 keV. Column 6 shows the count rate after vignetting correction. Column 7 shows the temperature for the blackbody fit to the spectrum ($kT_{\rm bb}$) in the range 0.3 to 8.0 keV. Column 8 shows the neutral hydrogen column density ($n_{\rm H}$). Column 9 shows the 0.3–8.0 keV flux for the blackbody fit to the spectrum. Column 10 shows the luminosity (L_X) at 0.3–8.0 keV. Column 11 shows the unabsorbed bolometric luminosity ($L_{\rm bol}$). Column 12 shows the reduced X^2 and degree of freedom (dof) for the spectral fit. Column 13 shows whether the observations indicate a high-flux or a low-flux state (10 counts per kilosecond separate these two states).

doi:10.1038/nature16067

Ab initio alpha–alpha scattering

Serdar Elhatisari[1], Dean Lee[2], Gautam Rupak[3], Evgeny Epelbaum[4], Hermann Krebs[4], Timo A. Lähde[5], Thomas Luu[1,5] & Ulf-G. Meißner[1,5,6]

Processes such as the scattering of alpha particles (^{4}He), the triple-alpha reaction, and alpha capture play a major role in stellar nucleosynthesis. In particular, alpha capture on carbon determines the ratio of carbon to oxygen during helium burning, and affects subsequent carbon, neon, oxygen, and silicon burning stages. It also substantially affects models of thermonuclear type Ia supernovae, owing to carbon detonation in accreting carbon–oxygen white-dwarf stars[1–3]. In these reactions, the accurate calculation of the elastic scattering of alpha particles and alpha-like nuclei—nuclei with even and equal numbers of protons and neutrons—is important for understanding background and resonant scattering contributions. First-principles calculations of processes involving alpha particles and alpha-like nuclei have so far been impractical, owing to the exponential growth of the number of computational operations with the number of particles. Here we describe an *ab initio* calculation of alpha–alpha scattering that uses lattice Monte Carlo simulations. We use lattice effective field theory to describe the low-energy interactions of protons and neutrons, and apply a technique called the 'adiabatic projection method' to reduce the eight-body system to a two-cluster system. We take advantage of the computational efficiency and the more favourable scaling with system size of auxiliary-field Monte Carlo simulations to compute an *ab initio* effective Hamiltonian for the two clusters. We find promising agreement between lattice results and experimental phase shifts for *s*-wave and *d*-wave scattering. The approximately quadratic scaling of computational operations with particle number suggests that it should be possible to compute alpha scattering and capture on carbon and oxygen in the near future. The methods described here can be applied to ultracold atomic few-body systems as well as to hadronic systems using lattice quantum chromodynamics to describe the interactions of quarks and gluons.

In recent years there has been much progress in *ab initio* scattering and reactions involving light[4–6] and medium-mass[7,8] nuclei. However, for most numerical methods, the number of computational operations increases markedly when the projectile nucleus has more than a few nucleons. Therefore it remains a challenge to study many important processes that are relevant for stellar astrophysics such as alpha–alpha scattering, alpha–carbon scattering and radiative capture, as well as carbon and oxygen burning in massive star evolution and thermonuclear supernovae[9].

We describe lattice calculations for which the number of computational (floating point) operations for the A_1-body + A_2-body problem scales as roughly $(A_1+A_2)^2$; this scaling is mild enough to make first-principles calculations of alpha processes possible. We use the formalism of lattice effective field theory[10–12] (EFT) and a technique for elastic scattering and inelastic reactions on the lattice called the 'adiabatic projection method'[13–17].

Chiral EFT is a framework for organizing the low-energy nuclear interactions of protons and neutrons according to powers of momenta and factors of the mass of the pion; see ref. 18 for a review of the theory. The important interactions are at leading order (LO), the next largest contributions are at next-to-leading order (NLO), and then follows next-to-next-to-leading order (NNLO). We present an *ab initio* calculation of ^{4}He + ^{4}He scattering going up to NNLO terms in chiral EFT. We find promising agreement with experimental data[19–22] for the *s*-wave and *d*-wave phase shifts; improvements can be achieved by including higher-order terms in the chiral expansion.

The adiabatic projection method addresses the cluster–cluster scattering problem on the lattice by using Euclidean time projection to construct an effective two-cluster Hamiltonian. By Euclidean time projection we mean multiplication by $\exp(-H\tau)$, where H is the underlying microscopic Hamiltonian and τ is Euclidean time. We use natural units, where the reduced Planck constant $\hbar$ and the speed of light c are set to one. Even though the actual lattice calculations use discrete time steps, we refer to the continuous Euclidean time parameter τ for notational simplicity.

Our starting point is a three-dimensional spatial lattice that is periodic with length L in each dimension. We take a set of initial two-alpha states $|\boldsymbol{R}\rangle$, labelled by their separation vector $\boldsymbol{R}$, as illustrated in Fig. 1. We take the initial alpha wavefunctions to be Gaussian wave packets, so that at large separations they factorize as a tensor product of two individual alpha clusters:

$$|\boldsymbol{R}\rangle = \sum_{\boldsymbol{r}} |\boldsymbol{r}+\boldsymbol{R}\rangle_1 \otimes |\boldsymbol{r}\rangle_2$$

where $\boldsymbol{r}$ is a summation variable corresponding to the location of the second cluster. The summation over $\boldsymbol{r}$ produces two-alpha states with total momentum equal to zero. Rather than dealing with a large array of three-dimensional vectors $\boldsymbol{R}$, we project onto spherical harmonics Y_{ℓ,ℓ_z} with angular momentum quantum numbers ℓ, ℓ_z:

$$|R\rangle^{\ell,\ell_z} = \sum_{\boldsymbol{R}'} Y_{\ell,\ell_z}(R')\delta_{R,|\boldsymbol{R}'|}|\boldsymbol{R}'\rangle$$

where δ is the Kronecker delta function. We only consider cases where $R = |\boldsymbol{R}| < L/2$.

On the lattice, the symmetry group of spatial rotations is broken down to a cubic subgroup. Nevertheless, at low scattering energies, this approximate rotational symmetry is very accurate, provided that artefacts due to the periodic volume are removed. We remove these artefacts using a hard spherical wall boundary; the spherical harmonic projection technique is useful for extracting data for selected partial waves. This method has been extended to particles with spin and partial wave mixing, and shows excellent agreement with continuous-space calculations[23].

We use Euclidean time projection to form dressed cluster states:

$$|R\rangle_\tau^{\ell,\ell_z} = \exp(-H\tau)|R\rangle^{\ell,\ell_z}$$

The evolution in Euclidean time automatically incorporates the induced deformation and polarization of the alpha clusters as they approach each other. The deformation and polarization are due to the

[1]Helmholtz-Institut für Strahlen- und Kernphysik and Bethe Center for Theoretical Physics, Universität Bonn, D-53115 Bonn, Germany. [2]Department of Physics, North Carolina State University, Raleigh, North Carolina 27695, USA. [3]Department of Physics and Astronomy and the HPC[2] Center for Computational Sciences, Mississippi State University, Mississippi State, Mississippi 39762, USA. [4]Institut für Theoretische Physik II, Ruhr-Universität Bochum, D-44870 Bochum, Germany. [5]Institute for Advanced Simulation, Institut für Kernphysik, and Jülich Center for Hadron Physics, Forschungszentrum Jülich, D-52425 Jülich, Germany. [6]JARA—High Performance Computing, Forschungszentrum Jülich, D-52425 Jülich, Germany.

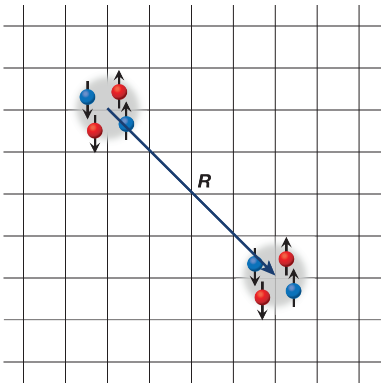


Figure 1 | Initial state clusters. Initial state $|\boldsymbol{R}\rangle$ composed of two alpha-particle wave packets on the lattice separated by the displacement vector $\boldsymbol{R}$. Each alpha-particle wave packet consists of four nucleons. Protons are red; neutrons are blue; spins are represented as arrows.

interactions of individual nucleons between the two alpha clusters, as well as to repulsion as a result of the Pauli exclusion principle for identical fermions.

With these dressed cluster states, we compute matrix elements of the full microscopic Hamiltonian with respect to the dressed cluster states:

$$[H_\tau]^{\ell,\ell_z}_{R,R'} = {}^{\ell,\ell_z}_{\tau}\langle R|H|R'\rangle^{\ell,\ell_z}_{\tau} \tag{1}$$

Because the dressed cluster states are not orthogonal, we construct a norm matrix:

$$[N_\tau]^{\ell,\ell_z}_{R,R'} = {}^{\ell,\ell_z}_{\tau}\langle R|R'\rangle^{\ell,\ell_z}_{\tau}$$

The radial adiabatic Hamiltonian is defined as a matrix product:

$$[H^a_\tau]^{\ell,\ell_z}_{R,R'} = [N_\tau^{-1/2}H_\tau N_\tau^{-1/2}]^{\ell,\ell_z}_{R,R'} \tag{2}$$

In the limit of large projection time τ, the spectrum of the adiabatic Hamiltonian reproduces the low-energy finite-volume spectrum of the microscopic Hamiltonian H. In ref. 17, it is shown that in the asymptotic region where the alpha clusters are widely separated, the adiabatic Hamiltonian reduces to a simple two-cluster Hamiltonian with only infinite-range interactions such as the Coulomb interaction between the otherwise non-interacting clusters. Although this may seem an obvious result, it is a non-trivial statement that the dependence on the projection time τ drops out from the adiabatic Hamiltonian at large distances.

We study ^{4}He + ^{4}He scattering using the same lattice action that is used to study the Hoyle state of ^{12}C (ref. 11). The spatial lattice spacing is $a = 1.97$ fm and the Euclidean-time, or temporal, lattice spacing is $a_t = 1.32$ fm. Revisiting these calculations in the future with different lattice spacings and including higher-order terms in the chiral expansion will provide a useful measure of systematic errors in lattice calculations of larger nuclear systems.

We perform projection Monte Carlo simulations with auxiliary fields to compute the matrices $[H_\tau]^{\ell,\ell_z}_{R,R'}$ and $[N_\tau]^{\ell,\ell_z}_{R,R'}$ on a periodic cubic lattice with volume $L^3 = (16\,\text{fm})^3$; see ref. 24 for an overview of methods used in lattice EFT. The total projection time for the initial and final dressed cluster states together is 2τ, which is equal to the product of the number of time steps L_t and the temporal lattice spacing a_t. We determine $[N_\tau]^{\ell,\ell_z}_{R,R'}$ from calculations with L_t time steps and $[H_\tau]^{\ell,\ell_z}_{R,R'}$ from calculations with $L_t + 1$ time steps. The extra time step for $[H_\tau]^{\ell,\ell_z}_{R,R'}$ is needed to calculate the matrix elements of H in equation (1). For these calculations, a new algorithm is used to allow for Monte Carlo updates of the auxiliary fields as well as updates of the alpha cluster positions.

We compute the radial adiabatic Hamiltonian using equation (2) and extend it to a much larger volume of $(120\,\text{fm})^3$. This is done by computing matrix elements of $[H^a_\tau]^{\ell,\ell_z}_{R,R'}$ at large separation (large R and R') from single-alpha lattice simulations, and then including the Coulomb interaction between the otherwise non-interacting clusters. This process also allows us to define a 'trivial' two-cluster Hamiltonian in which the two alpha clusters are non-interacting except for the infinite-range Coulomb interaction.

With the radial adiabatic Hamiltonian defined in the large $(120\,\text{fm})^3$ box, we extract the scattering phase shifts by imposing a hard spherical wall boundary at some radius R_{wall} and determining the standing wave modes. In Fig. 2 we show s-wave radial functions for two different radial excitations (2s and 3s) at NNLO using chiral EFT. The error bars show 1-standard deviation (s.d.) Monte Carlo errors calculated using a jackknife analysis of the lattice data. We could extract the phase shift by fitting to the asymptotic behaviour of the radial wavefunction as in ref. 17; however, it is more accurate to extract the phase shifts from the energy of the standing wave, as discussed in ref. 25.

Figure 3 shows the phase shifts for s-wave scattering versus laboratory energy at LO, NLO, and NNLO in chiral EFT, compared with experimental data[19–22]. The green dashed (LO), blue short-dashed (NLO), and red solid lines (NNLO) are determined from fits to the lattice data using the effective range expansion (see Methods). For further comparison, the inset of Fig. 3 shows NLO results using halo EFT with point-like alpha particles[26]. Halo EFT is an effective theory in which clusters of tightly bound nucleons are treated as point particles. Our LO results do not include Coulomb effects and so have substantially different behaviour near the alpha–alpha scattering threshold. The NLO and NNLO phase shifts are quite similar, and both agree fairly well with the experimental data. The close agreement between NLO and NNLO results is probably accidental: several contributions appearing at NNLO seem to cancel each other out. The same does not occur for the d-wave phase shifts. The results and error bars shown in Fig. 3 are computed from lattice phase-shift data for $L_t = 4$ to $L_t = 10$ and extrapolating to the limit $L_t \to \infty$. Details of the extrapolation fit and all associated error estimates are discussed in Methods. The observed energy of the s-wave resonance in the centre-of-mass frame is 0.09184 MeV above threshold. For the lattice results, we find that the ground state is 0.79(9) MeV below threshold at LO, and 0.11(1) MeV below threshold at both NLO and NNLO (the errors in parentheses here and elsewhere represent 1 s.d.).

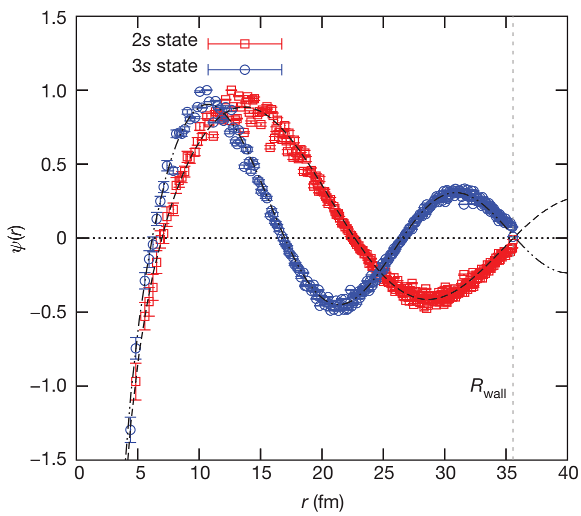


Figure 2 | s-wave scattering radial wavefunctions. The second-lowest-energy (red squares) and third-lowest-energy (blue circles) s-wave radial wavefunctions for spherical wall radius $R_{\text{wall}} \approx 36$ fm (grey dashed line) at NNLO plotted versus radial distance. The dashed and double-dot-dashed lines show the fits to a Coulomb wavefunction for the second and third radial states, respectively. The error bars indicate 1-s.d. Monte Carlo errors calculated using a jackknife analysis of the lattice data.

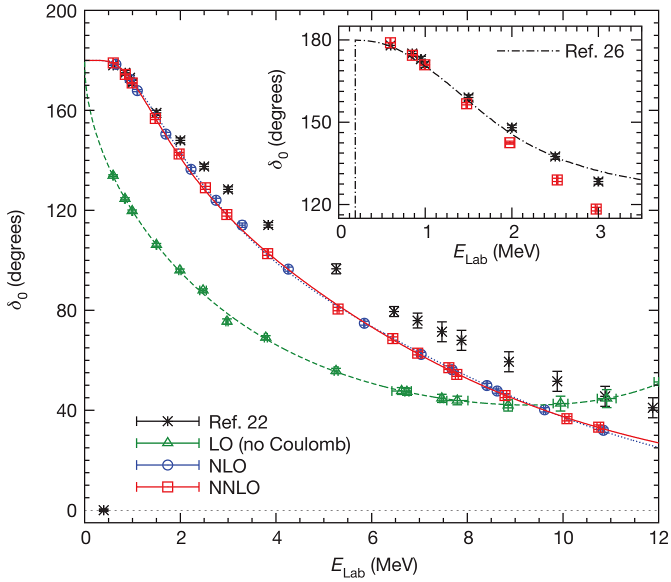


Figure 3 | *s*-wave phase shifts. *s*-wave phase shifts δ_0 at LO (green triangles), NLO (blue circles), and NNLO (red squares) versus laboratory energy E_{Lab}, compared with experimental data[19–22] (black asterisks). The theoretical error bars indicate 1 s.d. uncertainty due to Monte Carlo errors and the extrapolation of that data to infinite projection time. The green dashed (LO), blue short-dashed (NLO), and red solid (NNLO) lines are determined from fits to the lattice data using the effective range expansion. The black dot-dashed line in the inset shows NLO results using halo EFT with point-like alpha particles[26].

In Fig. 4 we show phase shifts for *d*-wave scattering versus laboratory energy at LO, NLO and NNLO, compared with experimental data[19–22]. The green dashed (LO), blue short-dashed (NLO), and red solid lines (NNLO) are determined from fits to the lattice data using the effective range expansion. Although there are differences, the NNLO results agree fairly with the experimental results. As in the *s*-wave case, we show the extrapolated values and errors in the limit $L_t \to \infty$, using lattice data for $L_t = 4$ to $L_t = 10$. Details of the extrapolation fit and all associated error estimates are discussed in Methods. We determined the centre-of-mass energy and the decay width of the *d*-wave resonance of the phase shift data from ref. 22 to be $E_R = 2.92(18)$ MeV and $\Gamma = 1.34(50)$ MeV, respectively. Owing to the large decay width, there is some model dependence in the definitions of the resonance parameters; we discuss several different definitions and determinations in Methods. At LO we find $E_R = 1.10(12)$ MeV and $\Gamma = 0.32(10)$ MeV, at NLO $E_R = 3.84(16)$ MeV and $\Gamma = 3.22(21)$ MeV, and at NNLO $E_R = 3.27(12)$ MeV and $\Gamma = 2.09(16)$ MeV.

To summarize, we present an *ab initio* calculation of ^{4}He + ^{4}He scattering. We use lattice EFT and the adiabatic projection method to compute phase shifts for *s*-wave and *d*-wave scattering up to NNLO, and find promising agreement with experimental data. To perform these calculations, we used spherical wave projections of the lattice initial states and a new algorithm that performs updates of both the auxiliary field configurations and alpha cluster positions. A schematic of the method is given in Extended Data Fig. 1.

Perhaps the most notable outcome of this study is a numerical method for simulating scattering and reactions that has a very favourable scaling with particle number. The number of computational operations needed for the A_1-body + A_2-body problem scales roughly as $(A_1 + A_2)^2$ for light and medium-mass nuclei, and the algorithm does not require the projectile to be very light. Because sign oscillations are greatly suppressed for alpha-like nuclei[12,27], our approach appears to be a viable method for studying important processes such as alpha scattering and capture on ^{12}C. Direct experimental data for alpha capture on ^{12}C is not possible, owing to Coulomb barrier suppression at energies relevant for stellar nucleosynthesis, and extrapolations from higher energies have uncertainties that exceed the 10% accuracy needed for stellar evolution models.

Nevertheless, there has been progress in measuring the contribution from subthreshold states[28] and cumulative *R*-matrix analyses using multiple data sources such as beta-delayed alpha-decay of ^{16}N and ^{4}He + ^{12}C elastic scattering[29]. *Ab initio* lattice calculations can contribute to these efforts by calculating asymptotic normalization coefficients for subthreshold states, determining the direct capture rate onto the ground state, and providing low-energy data on ^{4}He + ^{12}C elastic scattering. For these future calculations, we expect that about four times as much computing time as the roughly two million core hours used for this work will be required; the computational resources available appear sufficient to keep stochastic errors under control. To reduce systematic errors, we are currently working on including lattice nuclear forces at the next-higher order in the chiral expansion, reducing the lattice spacing, improving the lattice action, and doing precision tests of systematic errors in the adiabatic projection method. If necessary, the *ab initio* lattice results will be further improved by including short-range operators in the adiabatic Hamiltonian to make fine adjustments to the energies of near-threshold states of ^{16}O.

There is an obvious overlap between lattice calculations using the adiabatic projection method and halo EFT. Therefore it might be fruitful to look for synergies between the two methods. In cases where there is a large scale separation between the low-energy scattering and high-energy internal excitations, benchmark tests can be made between halo EFT and lattice calculations. Furthermore, *ab initio* calculations can be used to determine input data for halo EFT, as done in ref. 30. In cases where the separation of scales is not large, lattice calculations can be used to guide improvement of halo EFT to include nuclear core excitations. It also might be useful to treat the lattice adiabatic Hamiltonian as a halo EFT for clusters, and explore extensions to three- and four-cluster systems. This method could potentially be used to investigate multi-alpha-cluster structures in ^{12}C and ^{16}O.

It would be exciting to extend the methods presented here to lattice quantum chromodynamics (QCD) and construct adiabatic Hamiltonians for hadronic systems. All of the techniques used in our lattice simulations have immediate analogues in lattice QCD. The initial

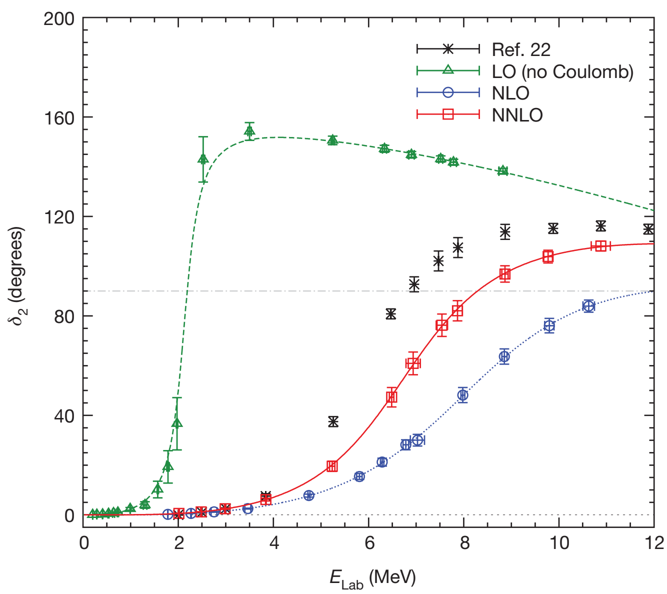


Figure 4 | *d*-wave phase shifts. *d*-wave phase shifts δ_2 at LO (green triangles), NLO (blue circles), and NNLO (red squares) versus laboratory energy E_{Lab}, compared with experimental data[19–22] (black asterisks). The theoretical error bars indicate 1 s.d. uncertainty due to Monte Carlo errors and the extrapolation of that data to infinite projection time. The green dashed (LO), blue short-dashed (NLO), and red solid (NNLO) lines are determined from fits to the lattice data using the effective range expansion.

cluster wavefunctions and stochastic sampling of cluster positions can be implemented using interpolating sources in lattice QCD. The lattice spherical harmonic projections introduced here and extended in ref. 23 should be useful for the treatment of angular momentum and partial waves in lattice QCD and constructing a hadronic adiabatic Hamiltonian.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 June; accepted 30 September 2015.

1. Imbriani, G. *et al.* The ^{12}C(α, γ)^{16}O reaction rate and the evolution of stars in the mass range $0.8 \leq M/M_{\odot} \leq 25$. *Astrophys. J.* **558,** 903–915 (2001).
2. Rauscher, T., Heger, A., Hoffman, R. D. & Woosley, S. E. Nucleosynthesis in massive stars with improved nuclear and stellar physics. *Astrophys. J.* **576,** 323–348 (2002).
3. Wiescher, M., Käppeler, F. & Langanke, K. Critical reactions in contemporary nuclear astrophysics. *Annu. Rev. Astron. Astrophys.* **50,** 165–210 (2012).
4. Nollett, K. M., Pieper, S. C., Wiringa, R. B., Carlson, J. & Hale, G. M. Quantum Monte Carlo calculations of neutron-α scattering. *Phys. Rev. Lett.* **99,** 022502 (2007).
5. Quaglioni, S. & Navrátil, P. *Ab initio* many-body calculations of n-^{3}H, n-^{4}He, p-^{3,4}He, and n-^{10}Be scattering. *Phys. Rev. Lett.* **101,** 092501 (2008).
6. Navrátil, P. & Quaglioni, S. *Ab initio* many-body calculations of the ^{3}H(d, n)^{4}He and ^{3}He(d, p)^{4}He fusion. *Phys. Rev. Lett.* **108,** 042503 (2012).
7. Hagen, G. & Michel, N. Elastic proton scattering of medium mass nuclei from coupled-cluster theory. *Phys. Rev. C* **86,** 021602(R) (2012).
8. Orlandini, G. *et al.* Coupling the Lorentz integral transform (LIT) and the coupled cluster (CC) methods: a way towards continuum spectra of “not-so-few-body” systems. *Few-Body Syst.* **55,** 907–911 (2014).
9. Woosley, S. E., Arnett, W. D. & Clayton, D. D. The explosive burning of oxygen and silicon. *Astrophys. J. Suppl. Ser.* **26,** 231–312 (1973).
10. Epelbaum, E., Krebs, H., Lee, D. & Meißner, U.-G. *Ab initio* calculation of the Hoyle state. *Phys. Rev. Lett.* **106,** 192501 (2011).
11. Epelbaum, E., Krebs, H., Lähde, T., Lee, D. & Meißner, U.-G. Structure and rotations of the Hoyle state. *Phys. Rev. Lett.* **109,** 252501 (2012).
12. Lähde, T. A. *et al.* Nuclear lattice simulations using symmetry-sign extrapolation. *Eur. Phys. J. A* **51,** 92 (2015).
13. Rupak, G. & Lee, D. Radiative capture reactions in lattice effective field theory. *Phys. Rev. Lett.* **111,** 032502 (2013).
14. Pine, M., Lee, D. & Rupak, G. Adiabatic projection method for scattering and reactions on the lattice. *Eur. Phys. J. A* **49,** 151 (2013).
15. Elhatisari, S. & Lee, D. Fermion-dimer scattering using an impurity lattice Monte Carlo approach and the adiabatic projection method. *Phys. Rev. C* **90,** 064001 (2014).
16. Rupak, G. & Ravi, P. Proton–proton fusion in lattice effective field theory. *Phys. Lett. B* **741,** 301–304 (2014).
17. Rokash, A. *et al.* Scattering cluster wave functions on the lattice using the adiabatic projection method. *Phys. Rev. C* **92,** 054612 (2015).
18. Epelbaum, E., Hammer, H.-W. & Meißner, U.-G. Modern theory of nuclear forces. *Rev. Mod. Phys.* **81,** 1773–1825 (2009).
19. Heydenburg, N. P. & Temmer, G. M. Alpha-alpha scattering at low energies. *Phys. Rev.* **104,** 123–134 (1956).
20. Nilson, R., Jentschke, W. K., Briggs, G. R., Kerman, R. O. & Snyder, J. N. Investigation of excited states in Be^{8} by alpha-particle scattering from He. *Phys. Rev.* **109,** 850–860 (1958).
21. Tombrello, T. A. & Senhouse, L. S. Elastic scattering of alpha particles from helium. *Phys. Rev.* **129,** 2252–2258 (1963).
22. Afzal, S. A., Ahmad, A. A. Z. & Ali, S. Systematic survey of the α–α interaction. *Rev. Mod. Phys.* **41,** 247–273 (1969).
23. Lu, B.-N., Lähde, T. A., Lee, D. & Meißner, U.-G. Precise determination of lattice phase shifts and mixing angles. Preprint at http://arxiv.org/abs/1506.05652 (2015).
24. Lee, D. Lattice simulations for few- and many-body systems. *Prog. Part. Nucl. Phys.* **63,** 117–154 (2009).
25. Borasoy, B., Epelbaum, E., Krebs, H., Lee, D. & Meißner, U.-G. Two-particle scattering on the lattice: phase shifts, spin-orbit coupling, and mixing angles. *Eur. Phys. J. A* **34,** 185–196 (2007).
26. Higa, R., Hammer, H.-W. & van Kolck, U. $\alpha\alpha$ scattering in halo effective field theory. *Nucl. Phys. A* **809,** 171–188 (2008).
27. Chen, J.-W., Lee, D. & Schaefer, T. Inequalities for light nuclei in the Wigner symmetry limit. *Phys. Rev. Lett.* **93,** 242302 (2004).
28. Avila, M. L. *et al.* Constraining the 6.05 MeV 0^{+} and 6.13 MeV 3^{-} cascade transitions in the ^{12}C(α, γ)^{16}O reaction using the asymptotic normalization coefficients. *Phys. Rev. Lett.* **114,** 071101 (2015).
29. Schürmann, D., Gialanella, L., Kunz, R. & Strieder, F. The astrophysical S factor of ^{12}C(α, γ)^{16}O at stellar energy. *Phys. Lett. B* **711,** 35–40 (2012).
30. Zhang, X., Nollett, K. M. & Phillips, D. R. Combining *ab initio* calculations and low-energy effective field theory for halo nuclear systems: the case of $^{7}\mathrm{Be} + p \to {}^{8}\mathrm{B} + \gamma$. *Phys. Rev. C* **89,** 051602(R) (2014).

Acknowledgements We acknowledge discussions with G. Hale and partial financial support from the Deutsche Forschungsgemeinschaft (Sino-German CRC 110), the Helmholtz Association (contract no. VH-VI-417), BMBF (grant no. 05P12PDFTE), the US Department of Energy (DE-FG02-03ER41260), and US National Science Foundation grant no. PHY-1307453. Further support was provided by the EU HadronPhysics3 project, the ERC project no. 259218 NUCLEAREFT, and the Magnus Ehrnrooth Foundation of the Finnish Society of Sciences and Letters. The computational resources were provided by the Jülich Supercomputing Centre at Forschungszentrum Jülich and by RWTH Aachen.

Author Contributions S.E. performed the analysis of the scattering data. S.E., E.E., H.K., D.L., and G.R. were involved in conceptual development of the adiabatic projection method. S.E. and T.L. produced the figures. T.A.L., D.L., T.L., and U.-G.M. obtained supercomputing time and developed the code. All authors were involved in writing, editing, and reviewing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.L. (dean_lee@ncsu.edu).

METHODS

Auxiliary field formalism. We simulate the interactions of nucleons on the lattice using projection Monte Carlo with auxiliary fields; see ref. 24 for an overview of methods used in lattice EFT. The details of the lattice action used in the calculations can be found in refs 11 and 31. We use an auxiliary-field formalism where the interactions among nucleons are replaced by interactions of nucleons with auxiliary fields at every lattice point in space and time[32–34]. This follows from an exact Gaussian integral identity connecting the exponential of the two-particle density ρ^2 to the integrated exponential of the one-particle density ρ:

$$\exp\left(-\frac{C}{2}\rho^2\right)=\sqrt{\frac{1}{2\pi}}\int_{-\infty}^{\infty}\mathrm{d}s\,\exp\left(-\frac{1}{2}s^2+\sqrt{-C}s\rho\right) \tag{3}$$

where C is the interaction coefficient and s is an auxiliary variable. In the auxiliary-field formalism each nucleon evolves as if it is a single particle in a fluctuating background of auxiliary fields. We use a total of sixteen auxiliary fields at LO in the chiral expansion coupled to the total nucleon density, three spin densities, three isospin densities, and nine spin–isospin densities. Similarly, the pion fields function much like auxiliary fields and generate the one-pion exchange potential. The interactions are reproduced by integrating over the auxiliary fields and pion fields. We use a spatial lattice spacing $a=(100\,\mathrm{MeV})^{-1}=1.97\,\mathrm{fm}$ and temporal lattice spacing $a_t=(150\,\mathrm{MeV})^{-1}=1.32\,\mathrm{fm}$.

For any fixed initial and final state, the amplitude for a given configuration of pion and auxiliary fields is proportional to the determinant of an $A\times A$ matrix M_{ij}. The entries of M_{ij} are the single nucleon amplitudes for a nucleon starting at state j at $\tau=0$ and ending at state i at $\tau=\tau_\mathrm{f}$. Formally, the calculation proceeds as follows. Let $|\Psi_{\boldsymbol{R}_1,\boldsymbol{R}_2}\rangle$ be an antisymmetrized product of single-nucleon states comprising two alpha clusters centred as Gaussian wave packets at each location $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$. Let H_LO denote the LO Hamiltonian that includes instantaneous one-pion exchange and contact interactions. Let $H_{\mathrm{SU}(4)}$ be an approximation to H_LO that has an underlying SU(4) symmetry among the nucleons that eliminates sign oscillations[27,35]. This SU(4) symmetry, discussed in ref. 36, refers to a symmetry group where the four nucleon states (proton spin-up, proton spin-down, neutron spin-up, neutron spin-down) can be interchanged with each other. This is a good starting point because the low-energy nuclear interactions are approximately SU(4) symmetric. The Coulomb interaction and the one-pion exchange interaction are certainly not SU(4) symmetric; however, the important short- and medium-range parts of the s-wave nucleon–nucleon interactions appear to respect this symmetry rather well[37]. There are arguments from quantum chromodynamics in the limit of a large number of colours that explain how this spin and flavour symmetry can arise[38,39]. There is also empirical evidence that some predictions of SU(4) symmetry are rather well satisfied by the spectrum of light nuclei[35], and lattice QCD simulations have shown that SU(4) symmetry becomes even more accurate for heavier quark masses[40,41].

Let us define a trial wavefunction:

$$|\Psi_{\boldsymbol{R}_1,\boldsymbol{R}_2}(\tau')\rangle=\exp(-H_{\mathrm{SU}(4)}\tau')|\Psi_{\boldsymbol{R}_1,\boldsymbol{R}_2}\rangle$$

We are using $\exp(-H_{\mathrm{SU}(4)}\tau')$ as an approximate low-energy filter that is computationally inexpensive. With this trial wavefunction we compute the amplitude:

$$Z(\boldsymbol{R}_3,\boldsymbol{R}_4;\boldsymbol{R}_1,\boldsymbol{R}_2;2\tau)=\langle\Psi_{\boldsymbol{R}_3,\boldsymbol{R}_4}(\tau')|\exp(-2H_\mathrm{LO}\tau)|\Psi_{\boldsymbol{R}_1,\boldsymbol{R}_2}(\tau')\rangle$$

We write 2τ rather than τ because the total projection time for the initial and final dressed cluster states together is 2τ. Higher-order contributions, Coulomb interactions, and isospin-breaking effects are computed as perturbative corrections to this amplitude. The auxiliary fields are updated using a non-local updating algorithm called hybrid Monte Carlo, while the coordinates of the alpha clusters in the initial and final states are updated using the Metropolis algorithm with random local updates.

Hybrid Monte Carlo algorithm. The hybrid Monte Carlo algorithm[42–44] is an efficient method for generating non-local updates of the auxiliary fields and pion fields. For simplicity we denote s as one of the sixteen possible auxiliary fields or three pion fields and discuss the updating algorithm for this field. The hybrid Monte Carlo algorithm was first introduced in lattice QCD, in which the size of the matrix is proportional to the number of space-time lattice points. In our case the matrix M_{ij} is an $A\times A$ matrix, where A is the number of nucleons. In general terms, the algorithm is described by means of a probability weight and a molecular-dynamics Hamiltonian:

$$P(s)\propto\exp[-V(s)],\quad H(s,p)\equiv\frac{1}{2}\sum_{\boldsymbol{n},n_\mathrm{t}}[p(\boldsymbol{n},n_\mathrm{t})]^2+V(s)$$

$V(s)$ is, in general, a non-local function of the field $s(\boldsymbol{n},n_\mathrm{t})$, where $\boldsymbol{n}$ denotes the spatial lattice site and n_t is the number of time steps; $p(\boldsymbol{n},n_\mathrm{t})$ denotes the momentum conjugate to $s(\boldsymbol{n},n_\mathrm{t})$. In our case $P(s)$ is a product of terms arising from the quadratic action for s (given in equation (3)) and the absolute value of the determinant of $M_{ij}(s)$.

Given an arbitrary initial configuration $s^0(\boldsymbol{n},n_\mathrm{t})$, the conjugate momentum is chosen from a random Gaussian distribution:

$$P[p^0(\boldsymbol{n},n_\mathrm{t})]\propto\exp\left\{-\frac{1}{2}[p^0(\boldsymbol{n},n_\mathrm{t})]^2\right\} \tag{4}$$

The Hamiltonian equations of motion are then integrated numerically with a finite step size $\varepsilon_\mathrm{step}$. We begin with a half-step forward in the conjugate momentum:

$$\tilde{p}^0(\boldsymbol{n},n_\mathrm{t})=p^0(\boldsymbol{n},n_\mathrm{t})-\frac{\varepsilon_\mathrm{step}}{2}\left[\frac{\partial V(s)}{\partial s(\boldsymbol{n},n_\mathrm{t})}\right]_{s=s^0}$$

followed by repeated updates of s and $\tilde{p}$:

$$s^{i+1}(\boldsymbol{n},n_\mathrm{t})=s^i(\boldsymbol{n},n_\mathrm{t})+\varepsilon_\mathrm{step}\tilde{p}^i(\boldsymbol{n},n_\mathrm{t}),$$
$$\tilde{p}^{i+1}(\boldsymbol{n},n_\mathrm{t})=\tilde{p}^i(\boldsymbol{n},n_\mathrm{t})-\varepsilon_\mathrm{step}\left[\frac{\partial V(s)}{\partial s(\boldsymbol{n},n_\mathrm{t})}\right]_{s=s^{i+1}}$$

for a specified number of steps N_step. The remaining half-step backward in $\tilde{p}$ is given by:

$$p^{N_\mathrm{step}}(\boldsymbol{n},n_\mathrm{t})=\tilde{p}^{N_\mathrm{step}}(\boldsymbol{n},n_\mathrm{t})+\frac{\varepsilon_\mathrm{step}}{2}\left[\frac{\partial V(s)}{\partial s(\boldsymbol{n},n_\mathrm{t})}\right]_{s=s^{N_\mathrm{step}}}$$

The evolved configuration is then subjected to a 'Metropolis test' using a random number $r\in[0,1)$. The condition:

$$r<\exp[-H(s^{N_\mathrm{step}},p^{N_\mathrm{step}})+H(s^0,p^0)]$$

determines whether the new configuration is accepted or rejected. If the Metropolis test is passed, then both s and p are updated, otherwise the original s is retained and only p is refreshed, according to equation (4).

Calculation of adiabatic projection matrix elements. Along with each hybrid Monte Carlo update, we allow for Metropolis updates of the locations of the Gaussian wave packets in the initial state, $\boldsymbol{R}_1$, $\boldsymbol{R}_2$ and the final state, $\boldsymbol{R}_3$, $\boldsymbol{R}_4$. This Metropolis update is accepted or rejected on the basis of the absolute value of the amplitude $Z(\boldsymbol{R}_3,\boldsymbol{R}_4;\boldsymbol{R}_1,\boldsymbol{R}_2;2\tau)$ for the new wave packet positions. We perform projection Monte Carlo simulations to compute the matrices $[H_\tau]^{\ell,\ell_z}_{R,R'}$ and $[N_\tau]^{\ell,\ell_z}_{R,R'}$ on a periodic cubic lattice with volume $L^3=(16\,\mathrm{fm})^3$ for partial waves $\ell=0$ and $\ell=2$. Separate simulations are performed for the s-wave and the d-wave. From the matrices $[H_\tau]^{\ell,\ell_z}_{R,R'}$ and $[N_\tau]^{\ell,\ell_z}_{R,R'}$, we compute the radial adiabatic Hamiltonian $[H^a_\tau]^{\ell,\ell_z}_{R,R'}$.

We consider values of the radial parameters R and R' ranging up to $L/3$. But we extend the radial Hamiltonian to much larger values of R and R' by computing a 'trivial' radial adiabatic Hamiltonian from single-alpha-cluster simulations and including the infinite-range Coulomb interaction explicitly. Let $|\Psi_{\boldsymbol{R}_1}\rangle$ be an antisymmetrized product of single-nucleon states comprising a single alpha cluster. As is done for the two-cluster simulations, we define a trial wavefunction:

$$|\Psi_{\boldsymbol{R}_1}(\tau')\rangle=\exp(-H_{\mathrm{SU}(4)}\tau')|\Psi_{\boldsymbol{R}_1}\rangle$$

With this trial wavefunction, we compute the LO one-cluster amplitude:

$$Z(\boldsymbol{R}_3;\boldsymbol{R}_1;2\tau)=\langle\Psi_{\boldsymbol{R}_3}(\tau')|\exp(-2H_\mathrm{LO}\tau)|\Psi_{\boldsymbol{R}_1}(\tau')\rangle$$

In the limit where $\boldsymbol{R}_1$ and $\boldsymbol{R}_3$ are widely separated from $\boldsymbol{R}_2$ and $\boldsymbol{R}_4$, we can factorize the LO two-cluster amplitude as a product of one-cluster amplitudes:

$$Z(\boldsymbol{R}_3,\boldsymbol{R}_4;\boldsymbol{R}_1,\boldsymbol{R}_2;2\tau)\rightarrow Z(\boldsymbol{R}_3;\boldsymbol{R}_1;2\tau)Z(\boldsymbol{R}_4;\boldsymbol{R}_2;2\tau)$$

In this manner we determine the trivial radial adiabatic Hamiltonian for large separation between the clusters. We use the trivial radial adiabatic Hamiltonian to extend $[H^a_\tau]^{\ell,\ell_z}_{R,R'}$ to volumes as large as $L^3=(120\,\mathrm{fm})^3$. This trivial radial adiabatic Hamiltonian does not take into account effects such as the antisymmetrization of identical nucleons from different clusters; however, such effects are expected to be negligible because we use the trivial Hamiltonian only at large distances where the clusters do not overlap. We performed simulations at a somewhat larger volume, $L^3=(20\,\mathrm{fm})^3$, and estimate that the error due to lack of full antisymmetrization of the trivial Hamiltonian results in errors smaller than a couple per cent for the phase shifts.

The NLO and NNLO corrections are included using first-order perturbation for the matrix $[H_\tau]^{\ell,\ell_z}_{R,R'}$, which yields a corresponding correction to the radial adiabatic Hamiltonian, $[H^a_\tau]^{\ell,\ell_z}_{R,R'}$. Even though we treat higher-order corrections to the

radial adiabatic Hamiltonian using perturbation theory, the scattering states of the radial adiabatic Hamiltonian are determined by imposing a hard spherical wall at radius R_{wall} and using sparse eigenvector methods to determine the standing waves[25,45]. Therefore, at NLO and higher, we recover important non-perturbative Coulomb interactions near threshold. More detailed discussions of the infrared enhancement of Coulomb effects and strategies for power counting in EFT are found in refs 46 and 47.

The Coulomb interaction is included using perturbation theory in the lattice Monte Carlo simulations because a non-pertubative treatment to all orders would produce sign oscillations that diminish the quality of the data. For these calculations, however, there is no need to treat the Coulomb interaction non-perturbatively in the Monte Carlo simulations. The correction to the alpha binding energy due to the Coulomb interaction is less than 10%, and so higher-order Coulomb effects are at the level of 1% or smaller. The Coulomb interaction becomes important only at long distances, owing to the different asymptotic properties of Coulomb wavefunctions and spherical Bessel functions. This difference between Coulomb wavefunctions and spherical Bessel functions is properly handled in our calculations by the non-perturbative treatment of the Coulomb interaction in the adiabatic Hamiltonian.

For alpha-like nuclei, with even and equal numbers of protons and neutrons, sign oscillations are suppressed, owing to the approximate SU(4) symmetry of the low-energy interactions[12,27,35,36]. For very light nuclei, the computing time is dominated by the calculation of the single nucleon amplitudes, which scales linearly with the number of nucleons A. For very heavy nuclei, the computational time is mostly consumed by calculations of matrix determinants, which scale as A^3. For light to medium-mass nuclei, the scaling of the number of computational operations is roughly A^2, which is between A and A^3. The simulations were run with up to 32,768 cores on the JUQUEEN Blue Gene/Q supercomputer at the Jülich Supercomputing Centre. Roughly two million core hours were needed for the calculations reported here.

Effective range expansion. The Coulomb interaction does not appear in the LO term in the chiral expansion because we consider the electric charge e as a small parameter. There are alternative schemes that take into account enhancement of the Coulomb interaction near threshold[46,47]. We use the effective range expansion[48] to draw the green dashed lines through the LO data points in Figs 3 and 4. For partial wave ℓ, we fit the generalized scattering length a_ℓ, effective range r_ℓ, and shape parameter P_ℓ using:

$$p^{2\ell+1}\cot[\delta_\ell(p)] = -\frac{1}{a_\ell} + \frac{1}{2}r_\ell p^2 + P_\ell p^4 + \cdots \tag{5}$$

where p is the relative momentum.

For the NLO and NNLO results, we use the Coulomb-modified effective range expansions to draw the blue short-dashed (NLO) and red solid (NNLO) lines in Figs 3 and 4. We now briefly outline the elements of the Coulomb-modified effective range expansion here. Let us define the Coulomb parameter:

$$\gamma = 2\mu\alpha_{\mathrm{EM}}Z_1Z_2$$

where μ equals the reduced mass of the two-alpha system, $\alpha_{\mathrm{EM}} \approx 1/137$ is the electromagnetic fine-structure constant, and $Z_1 = Z_2 = 2$ are the charges of the two alpha particles. The factor $C^2_{\eta,\ell}$ for partial wave ℓ is defined as:

$$C^2_{\eta,\ell} = \frac{2^{2\ell}}{[(2\ell+1)!]^2}C^2_{\eta,0}\prod_{s=1}^{\ell}(s^2+\eta^2)$$

where

$$C^2_{\eta,0} = \frac{2\pi\eta}{e^{2\pi\eta}-1}$$

and $\eta = \gamma/(2p)$. The Coulomb-modified effective range expansion[48–50] is then:

$$C^2_{\eta,\ell}p^{2\ell+1}\cot[\delta_\ell(p)] + \gamma h_\ell(p) = -\frac{1}{a_\ell} + \frac{1}{2}r_\ell p^2 + P_\ell p^4 + \cdots \tag{6}$$

where:

$$h_\ell(p) = p^{2\ell}\frac{C^2_{\eta,\ell}}{C^2_{\eta,0}}h(\eta), \quad h(\eta) = \mathrm{Re}[\psi(i\eta)] - \log|\eta|$$

and $\psi(z) = \Gamma'(z)/\Gamma(z)$, in which the prime indicates differentiation.

We used the effective range expansion (equation (5)) and Coulomb-modified effective range expansion (equation (6)) to extract s-wave and d-wave energy levels. Although there is no problem extracting energy levels from the lattice data, the experimentally determined alpha–alpha s-wave phase shifts are not of sufficient quality near threshold to accurately determine the s-wave resonance parameters, and so they must be determined from other reactions. The Triangle Universities Nuclear Laboratory (TUNL) nuclear data evaluation gives the values[51] $E_{\mathrm{R}} = 0.09184$ MeV and $\Gamma = 5.57(25)$ eV.

The d-wave resonance extraction must be treated with some caution because the large decay width results in considerable model dependence, owing to different definitions of the resonance parameters and different fits to experimental data[52,53]. Here we use the definition that the resonance energy E_{R} is given by the location of the maximum of $\mathrm{d}\delta/\mathrm{d}E$, and the decay width Γ is determined by the value of $2(\mathrm{d}\delta/\mathrm{d}E)^{-1}$ at E_{R} (ref. 54). Using this definition, the d-wave phase shifts given in ref. 22 yield $E_{\mathrm{R}} = 2.92(18)$ MeV and $\Gamma = 1.34(50)$ MeV. Using the set of phase shifts provided in ref. 55, we obtain $E_{\mathrm{R}} = 2.88(6)$ MeV and $\Gamma = 1.52(3)$ MeV. The TUNL nuclear data evaluation reports values of $E_{\mathrm{R}} = 3.12(1)$ MeV and $\Gamma = 1.513(15)$ MeV.

Nuclear forces at NNLO. In the chiral nuclear EFT used here, the interactions among nucleons are organized according to their importance on the basis of a systematic expansion in powers of Q/Λ, with the 'hard scale' of the nuclear interactions $\Lambda \approx 1$ GeV. The 'soft scale' Q is associated with nucleon spatial momenta and the pion mass M_π. The dominant contributions to the nuclear Hamiltonian appear at $\mathcal{O}((Q/\Lambda)^0)$ (LO); the NLO terms are $\mathcal{O}((Q/\Lambda)^2)$ and involve only the two-nucleon force. In the results presented here, contributions to the nuclear Hamiltonian are taken into account up to $\mathcal{O}((Q/\Lambda)^3)$ (NNLO). These terms include the three-nucleon force, which first appears at NNLO. The electromagnetic force, which is important in nuclear binding, is also included consistently and systematically (for details, see ref. 31). At LO and NLO, we have two and seven low-energy constants, respectively. These constants are determined from a fit to neutron–proton scattering data. In addition, two low-energy constants parameterize the breaking of isospin symmetry of the strong nuclear force and are fixed from the proton–proton and neutron–neutron scattering lengths. Isospin symmetry refers to the equivalence of protons and neutrons and is an approximate symmetry of the nuclear interactions. There are also two low-energy constants parameterizing the three-nucleon force that are determined from the triton binding energy and the axial vector current contribution to triton decay[56].

Data extrapolation and error analysis. We use the lattice Monte Carlo data to determine the radial adiabatic Hamiltonian for the s-wave and d-wave channels for $L_t = 4$ to $L_t = 10$. We then compute the s-wave and d-wave phase shifts with errors calculated using a jackknife analysis of the Monte Carlo data. In Extended Data Figs 2 and 3 we show NNLO results for the s-wave and d-wave phase shifts, respectively.

The dot-dashed lines in Extended Data Figs 2 and 3 indicate the fitted exponential curves that are used to extrapolate to the limit $L_t \to \infty$. This is done by including the residual dependence from an excited state at energy ΔE above the ground-state energy. We use the ansatz:

$$\delta_0(L_t, E) = \delta_0(E) + c_0(E)\exp[-\Delta E_0 L_t a_t]$$

for the s-wave and:

$$\delta_2(L_t, E) = \delta_2(E) + c_2(E)\exp[-\Delta E_2 L_t a_t]$$

for the d-wave, where $c_0(E)$ and $c_2(E)$ are fit parameters; see ref. 17 for a discussion of the asymptotic time dependence of the adiabatic projection method. The dependence on L_t is caused by a residual contamination due to excited states other than alpha–alpha scattering states. We expect the convergence in L_t to be quite fast because there is a rather large energy gap between these excited states and the alpha–alpha scattering threshold. This fast convergence is seen in Extended Data Figs 2 and 3. The hatched bands show the one standard deviation errors of the extrapolation fits, including the propagated Monte Carlo errors of the data points.

Dependence on lattice spacing. The simulations reported here were performed on a coarse lattice with a lattice spacing $a = 1.97$ fm. This raises the question of whether there are sizeable lattice artefacts related to this spacing. We are currently working on simulations of lattice chiral EFT with smaller lattice spacings and improved lattice actions, which remove lattice artefacts to higher orders. Studies of lattice spacing dependence for the two-nucleon system[57] and for the alpha–alpha system with point-like alpha particles[58,59] found that observables that probe low-energy phenomena are largely independent of the lattice spacing in the interval $a = 0.5$–2 fm; strategies to further reduce lattice artefacts are discussed in refs 58 and 59. Hence, we expect little dependence on lattice spacing for low-energy scattering between two alpha particles, provided that the binding energy and structure of the alpha particles are well reproduced, as they are for the lattice action used here.

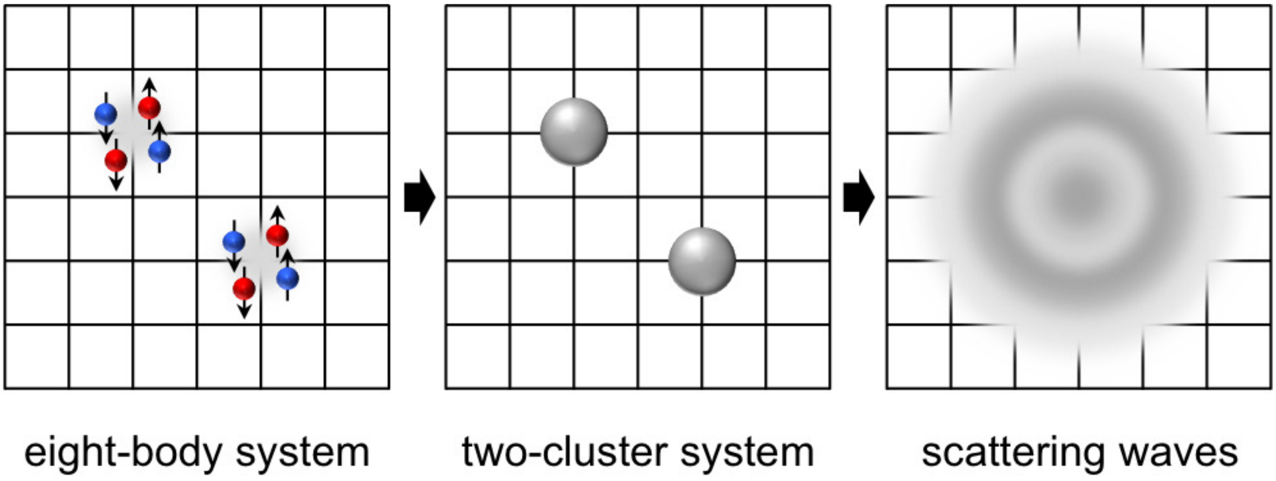
Future prospects for lattice simulations in nuclear physics. Lattice simulations have been increasingly important for the development of *ab initio* nuclear theory. Starting from a theory for quarks and gluons, lattice QCD has enabled promising steps towards calculating the interactions of nucleons and very light nuclei[40,60,61], as well as of magnetic moments[62] and radiative capture[41]. Starting from a theory for protons and neutrons, lattice EFT has enabled promising steps towards calculating

the structure of light and medium-mass nuclei[10,11,63,64] and neutron matter[65,66], and is now allowing a move towards calculating nuclear scattering and reactions. The two methods complement one another, with lattice QCD providing input data for lattice EFT, and lattice EFT providing a computationally efficient method of simulating systems with many nucleons. The transfer of data can be realized by tuning operator coefficients in lattice EFT to match finite-volume energy levels of few-nucleon systems computed in lattice QCD simulations. Future calculations could start from quark masses and the electromagnetic fine-structure constant in lattice QCD, and end with reactions involving medium-mass nuclei in lattice EFT.

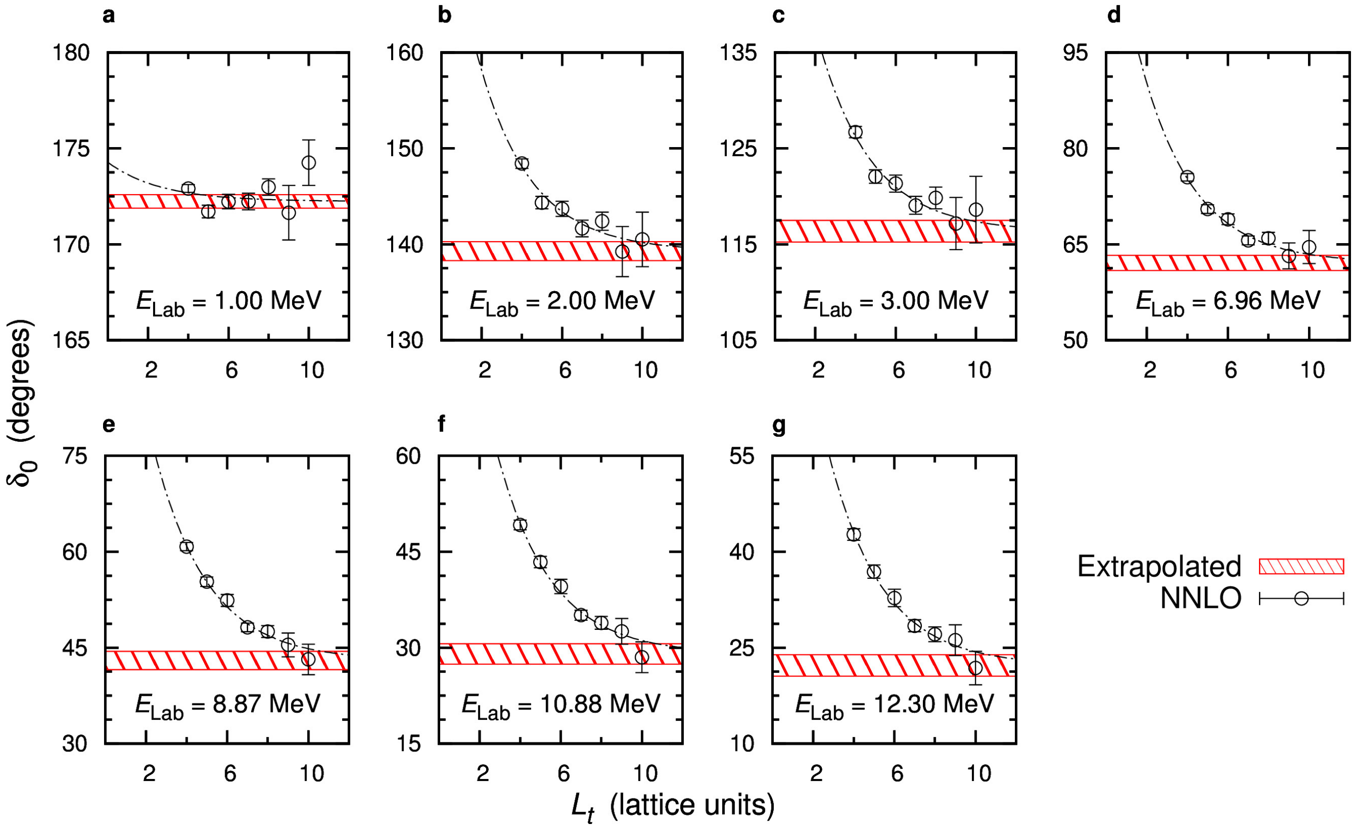
Additional references. Other work on halo EFT includes studies on ^{7}Li (ref. 67), ^{19}C (ref. 68), ^{6}He (ref. 69), and ^{7}Be (ref. 70).

Code availability. All codes used in this work are freely available from the authors on request.

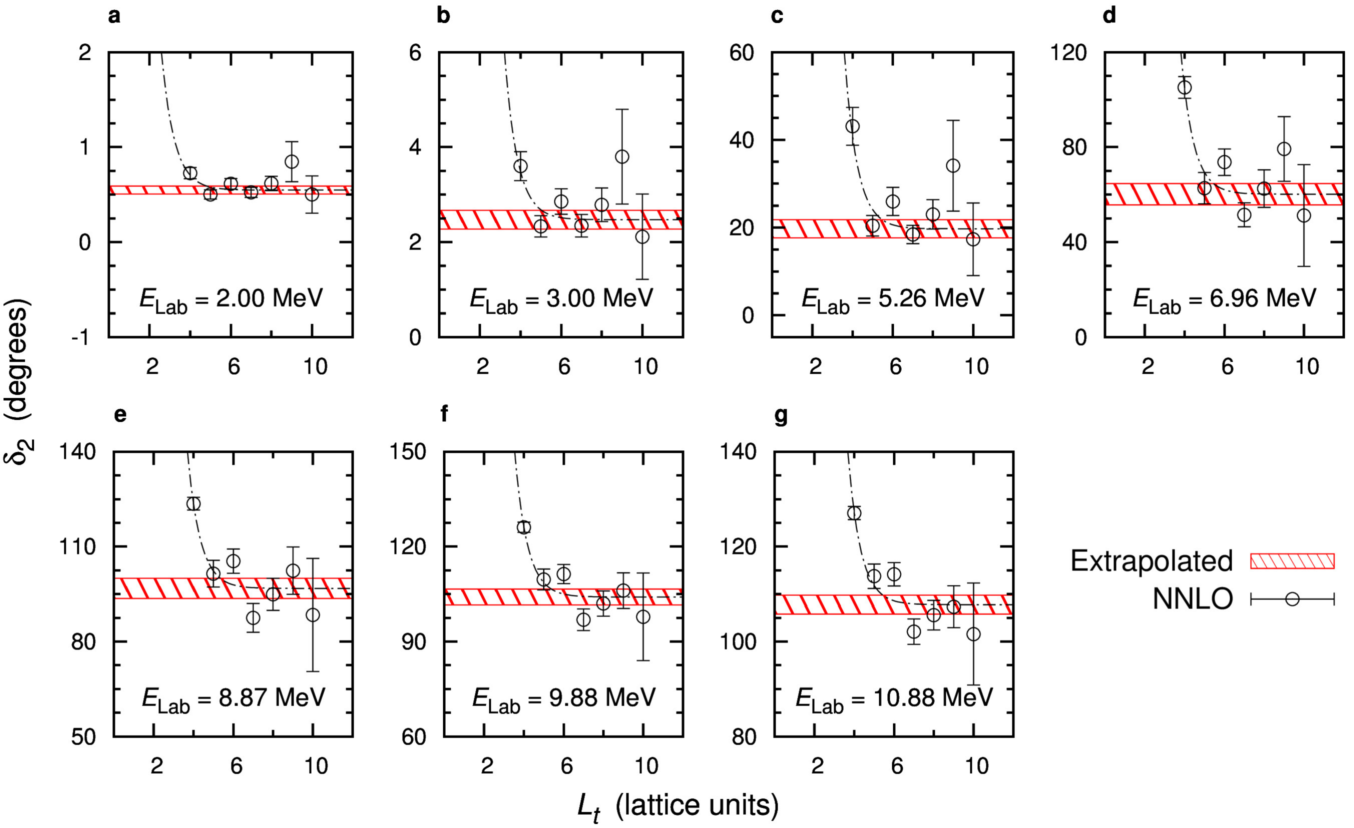
31. Epelbaum, E., Krebs, H., Lee, D. & Meißner, U.-G. Lattice calculations for A = 3, 4, 6, 12 nuclei using chiral effective field theory. *Eur. Phys. J. A* **45,** 335–352 (2010).
32. Hubbard, J. Calculation of partition functions. *Phys. Rev. Lett.* **3,** 77–78 (1959).
33. Stratonovich, R. L. On a method of calculating quantum distribution functions. *Sov. Phys. Dokl.* **2,** 416–419 (1958).
34. Koonin, S. E. Auxiliary-field Monte Carlo methods. *J. Stat. Phys.* **43,** 985–990 (1986).
35. Lee, D. Spectral convexity for attractive SU(2*N*) fermions. *Phys. Rev. Lett.* **98,** 182501 (2007).
36. Wigner, E. On the consequences of the symmetry of the nuclear hamiltonian on the spectroscopy of nuclei. *Phys. Rev.* **51,** 106–119 (1937).
37. Mehen, T., Stewart, I. W. & Wise, M. B. Wigner symmetry in the limit of large scattering lengths. *Phys. Rev. Lett.* **83,** 931–934 (1999).
38. Kaplan, D. B. & Savage, M. J. The spin-flavor dependence of nuclear forces from large-*N* QCD. *Phys. Lett. B* **365,** 244–251 (1996).
39. Calle Cordón, A. & Ruiz Arriola, E. Wigner symmetry, large N_c and renormalized one-boson exchange potential. *Phys. Rev. C* **78,** 054002 (2008).
40. Beane, S. *et al.* Nucleon-nucleon scattering parameters in the limit of SU(3) flavor symmetry. *Phys. Rev. C* **88,** 024003 (2013).
41. Beane, S. R. *et al. Ab initio* calculation of the $np \to d\gamma$ radiative capture process. *Phys. Rev. Lett.* **115,** 132001 (2015).
42. Scalettar, R. T., Scalapino, D. J. & Sugar, R. L. New algorithm for the numerical simulation of fermions. *Phys. Rev. B* **34,** 7911–7917 (1986).
43. Gottlieb, S., Liu, W., Toussaint, D., Renken, R. L. & Sugar, R. L. Hybrid molecular dynamics algorithms for the numerical simulation of quantum chromodynamics. *Phys. Rev. D* **35,** 2531–2542 (1987).
44. Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. Hybrid Monte Carlo. *Phys. Lett. B* **195,** 216–222 (1987).
45. Carlson, J., Pandharipande, V. & Wiringa, R. Variational calculations of resonant states in ^{4}He. *Nucl. Phys. A* **424,** 47–59 (1984).
46. Kong, X. & Ravndal, F. Coulomb effects in low-energy proton proton scattering. *Nucl. Phys. A* **665,** 137–163 (2000).
47. Rupak, G. & Kong, X.-w. Quartet *S*-wave *p*–*d* scattering in EFT. *Nucl. Phys. A* **717,** 73–90 (2003).
48. Bethe, H. A. Theory of the effective range in nuclear scattering. *Phys. Rev.* **76,** 38–50 (1949).
49. Jackson, J. D. & Blatt, J. M. The interpretation of low energy proton-proton scattering. *Rev. Mod. Phys.* **22,** 77–118 (1950).
50. König, S., Lee, D. & Hammer, H.-W. Causality constraints for charged particles. *J. Phys. G* **40,** 045106 (2013).
51. Tilley, D. R. *et al.* Energy levels of light nuclei A = 8, 9, 10. *Nucl. Phys. A* **745,** 155–362 (2004).
52. Hoop, B., Hale, G. M. & Navratil, P. Neutron-^{4}He resonant scattering at d-^{3}H threshold. Preprint at http://arxiv.org/abs/1111.0985 (2011).
53. Orlov, Yu. V., Irgaziev, B. F. & Nikitina, L. I. Asymptotic normalization coefficients of resonant and bound states from the phase shifts for $\alpha\alpha$ and α^{12}C scattering. Preprint at http://arxiv.org/abs/1508.07538 (2015).
54. Hupin, G., Quaglioni, S. & Navrátil, P. Predictive theory for elastic scattering and recoil of protons from 4He. *Phys. Rev. C* **90,** 061601(R) (2014).
55. Page, P. R. & Hale, G. M. ^{8}Be nuclear data evaluation. *AIP Conf. Proc.* **769,** 390–393 (2005).
56. Gazit, D., Quaglioni, S. & Navrátil, P. Three-nucleon low-energy constants from the consistency of interactions and currents in chiral effective field theory. *Phys. Rev. Lett.* **103,** 102502 (2009).
57. Klein, N., Lee, D., Liu, W. & Meißner, U.-G. Regularization methods for nuclear lattice effective field theory. *Phys. Lett. B* **747,** 511–516 (2015).
58. Lu, B.-N., Lähde, T. A., Lee, D. & Meißner, U.-G. Breaking and restoration of rotational symmetry on the lattice for bound state multiplets. *Phys. Rev. D* **90,** 034507 (2014).
59. Lu, B.-N., Lähde, T. A., Lee, D. & Meißner, U.-G. Breaking and restoration of rotational symmetry for irreducible tensor operators on the lattice. *Phys. Rev. D* **92,** 014506 (2015).
60. Yamazaki, T., Ishikawa, K.-i., Kuramashi, Y. & Ukawa, A. Helium nuclei, deuteron and dineutron in 2 + 1 flavor lattice QCD. *Phys. Rev. D* **86,** 074514 (2012).
61. Berkowitz, E. *et al.* Two-nucleon higher partial-wave scattering from lattice QCD. Preprint at http://arxiv.org/abs/1508.00886 (2015).
62. Chang, E. *et al.* The magnetic structure of light nuclei from lattice QCD. Preprint at http://arxiv.org/abs/1506.05518 (2015).
63. Epelbaum, E., Krebs, H., Lee, D. & Meißner, U.-G. Lattice effective field theory calculations for A = 3, 4, 6, 12 nuclei. *Phys. Rev. Lett.* **104,** 142501 (2010).
64. Lähde, T. A. *et al.* Lattice effective field theory for medium-mass nuclei. *Phys. Lett. B* **732,** 110–115 (2014).
65. Epelbaum, E., Krebs, H., Lee, D. & Meißner, U.-G. Ground state energy of dilute neutron matter at next-to-leading order in lattice chiral effective field theory. *Eur. Phys. J. A* **40,** 199–213 (2009).
66. Wlazłowski, G., Holt, J. W., Moroz, S., Bulgac, A. & Roche, K. J. Auxiliary-field quantum Monte Carlo simulations of neutron matter in chiral effective field theory. *Phys. Rev. Lett.* **113,** 182503 (2014).
67. Rupak, G. & Higa, R. Model-independent calculation of radiative neutron capture on lithium-7. *Phys. Rev. Lett.* **106,** 222501 (2011).
68. Acharya, B. & Phillips, D. R. ^{19}Carbon in halo EFT: effective-range parameters from Coulomb-dissociation experiments. *Nucl. Phys. A* **913,** 103–115 (2013).
69. Ji, C., Elster, C. & Phillips, D. R. ^{6}He nucleus in halo effective field theory. *Phys. Rev. C* **90,** 044004 (2014).
70. Ryberg, E., Forssén, C., Hammer, H.-W. & Platter, L. Constraining low-energy proton capture on beryllium-7 through charge radius measurements. *Eur. Phys. J. A* **50,** 170 (2014).



Extended Data Figure 1 | Schematic overview of our method. We start with an eight-body system of protons and neutrons. Each alpha-particle wave packet consists of four nucleons. The protons are red, the neutrons are blue, and the spins are represented by arrows. Next, we perfom *ab initio* lattice Monte Carlo simulations to construct the adiabatic Hamiltonian for two alpha clusters (grey spheres). Finally, we use the adiabatic Hamiltonian to compute alpha–alpha scattering phase shifts.



Extended Data Figure 2 | *s*-wave extrapolations at NNLO. a–g, NNLO results (circles) for the *s*-wave phase shift δ_0 versus L_t at laboratory energy $E_{\mathrm{Lab}} = 1.00$ MeV, 2.00 MeV, 3.00 MeV, 6.96 MeV, 8.87 MeV, 10.88 MeV, and 12.30 MeV, respectively, as labelled. The theoretical error bars indicate 1 s.d. uncertainty due to Monte Carlo errors. The dot-dashed lines are fits to the data, used to extrapolate the $L_t \to \infty$ limits. The red hatched regions indicate the 1 s.d. error estimate of the extrapolation.



Extended Data Figure 3 | *d*-wave extrapolations at NNLO. a–g, NNLO results (circles) for the *d*-wave phase shift δ_2 versus L_t at laboratory energy $E_{\mathrm{Lab}} = 2.00$ MeV, 3.00 MeV, 5.26 MeV, 6.96 MeV, 8.87 MeV, 9.88 MeV, and 10.88 MeV, respectively, as labelled. The theoretical error bars indicate 1 s.d. uncertainty due to Monte Carlo errors. The dot-dashed lines are fits to the data, used to extrapolate the $L_t \to \infty$ limits. The red hatched regions indicate the 1 s.d. error estimate of the extrapolation.

doi:10.1038/nature16147

Potential sea-level rise from Antarctic ice-sheet instability constrained by observations

Catherine Ritz[1,2]*, Tamsin L. Edwards[3,4]*, Gaël Durand[1,2], Antony J. Payne[4], Vincent Peyaud[1,2] & Richard C. A. Hindmarsh[5]

Large parts of the Antarctic ice sheet lying on bedrock below sea level may be vulnerable to marine-ice-sheet instability (MISI)[1], a self-sustaining retreat of the grounding line triggered by oceanic or atmospheric changes. There is growing evidence[2–4] that MISI may be underway throughout the Amundsen Sea embayment (ASE), which contains ice equivalent to more than a metre of global sea-level rise. If triggered in other regions[5–8], the centennial to millennial contribution could be several metres. Physically plausible projections are challenging[9]: numerical models with sufficient spatial resolution to simulate grounding-line processes have been too computationally expensive[2,3,10] to generate large ensembles for uncertainty assessment, and lower-resolution model projections[11] rely on parameterizations that are only loosely constrained by present day changes. Here we project that the Antarctic ice sheet will contribute up to 30 cm sea-level equivalent by 2100 and 72 cm by 2200 (95% quantiles) where the ASE dominates. Our process-based, statistical approach gives skewed and complex probability distributions (single mode, 10 cm, at 2100; two modes, 49 cm and 6 cm, at 2200). The dependence of sliding on basal friction is a key unknown: nonlinear relationships favour higher contributions. Results are conditional on assessments of MISI risk on the basis of projected triggers under the climate scenario A1B (ref. 9), although sensitivity to these is limited by theoretical and topographical constraints on the rate and extent of ice loss. We find that contributions are restricted by a combination of these constraints, calibration with success in simulating observed ASE losses, and low assessed risk in some basins. Our assessment suggests that upper-bound estimates from low-resolution models and physical arguments[9] (up to a metre by 2100 and around one and a half by 2200) are implausible under current understanding of physical mechanisms and potential triggers.

It is not yet clear[9] whether human-induced climate change has influenced the circulation of warm Circumpolar Deep Water driving grounding-line retreat[4] of Pine Island Glacier, Thwaites Glacier and other glaciers in the ASE, or how this circulation might change in future[9]. However, grounding-line retreat under MISI is proposed to occur at a rate more or less independent of the original trigger and may continue even if that trigger diminishes[2]. MISI can be limited by buttressing from ice shelves or specific configurations of bedrock topography[1,12] and possibly also higher friction at the bed[2,13,14]. It has been suggested that grounding-line retreat could continue in the ASE for decades[2] to centuries[3,4] owing to weak topographical constraints, possibly slowed in Pine Island Glacier by a region of higher friction behind the grounding line[2,13,14]. MISI could be triggered elsewhere by ice-shelf collapse and/or exposure of further ice shelves to Circumpolar Deep Water, both of which are projected in some regions[6,7] under the Special Report on Emissions Scenarios (SRES) A1B climate scenario[9]. Here we aim to quantify the dynamic contribution of the Antarctic ice sheet to sea level in the event of MISI under A1B.

We take a statistical–physical approach, using a numerical ice-sheet model[15] supplemented by statistical modelling of the probability of MISI onset. The statistical modelling represents the ocean and atmospheric drivers of MISI and the response of ice shelves, which are poorly known owing to the modelling challenges described earlier. We assign probabilities of MISI onset as a function of time until 2200 in each of 11 sectors (Extended Data Fig. 1a) using expert synthesis of observed grounding-line retreat and thinning[4,16,17] and projected ice-shelf basal[6,18] and surface[7] melting under A1B.

The response of the grounding-line position to MISI onset is represented with a new parameterization: if a MISI trigger occurs in a sector, the potential rate of retreat is a function of the basal friction coefficient at each part of the current grounding line (Extended Data Fig. 2c–e), with the form of the dependence (Extended Data Fig. 1b) based on theoretical considerations[1]. Grounding-line response is modified by two ice dynamical conditions that allow retreat to occur only if bedrock is downsloping from the margin (but allowing retreat over small bumps) and only at a rate not exceeding the theoretical limit[1]. The response is also modified by the basal friction law—the relationship between basal friction and sliding velocity—which has three possible configurations in this study: linear–viscous, nonlinear Weertman, or plastic flow.

To assess modelling uncertainties, we generated a 3,000-member ensemble sampling MISI onset dates in the 11 sectors, 3 parameters governing retreat rate, bedrock topography, and the form of the basal friction law. We weighted the ensemble members in a Bayesian statistical framework with the difference between simulated and observed mass losses in the ASE (the only region where grounding-line retreat has been observed) to obtain calibrated projections. Details and projections are in Supplementary Information.

Observational calibration gives greatest weight to the ensemble members that most successfully simulate present day ASE mass loss. The expected mass trend from 1992 to 2011 is $-59.0 \pm 13.5\,\mathrm{Gt\,yr^{-1}}$, where the standard deviation is dominated by a conservative tolerance for model error (Supplementary Information, section 1.7). The range of simulated mass trends is -13.4 to $-218.3\,\mathrm{Gt\,yr^{-1}}$, with 39% of the ensemble more than three standard deviations from the expected trend, of which nearly all simulate losses that are too large. Parameter values that generate the most rapid and widespread present day retreat in the ASE are thus effectively ruled out. These also tend to give the highest sea-level projections, so calibration decreases projected quantiles. Medians at 2100 and 2200 decrease by 33% and 20%, and 95% quantiles by 36% and 30%, respectively; the modes, however, increase, particularly at 2200 owing to a shift in density from one local mode to the other.

Spatial patterns of the probability of ungrounding (Fig. 1) show how local bed elevation, slope and friction strongly modulate the response to MISI onset. We find that the region with the highest probability of ungrounding and sea-level contribution is the ASE, owing to the

[1]CNRS, LGGE, F-38041 Grenoble, France. [2]Université Grenoble Alpes, LGGE, F-38041 Grenoble, France. [3]Department of Environment, Earth and Ecosystems, Faculty of Science, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK. [4]Department of Geographical Sciences, University of Bristol, University Road, Bristol BS8 1SS, UK. [5]British Antarctic Survey, Natural Environment Research Council, Madingley Road, Cambridge CB3 0ET, UK.
*These authors contributed equally to this work.

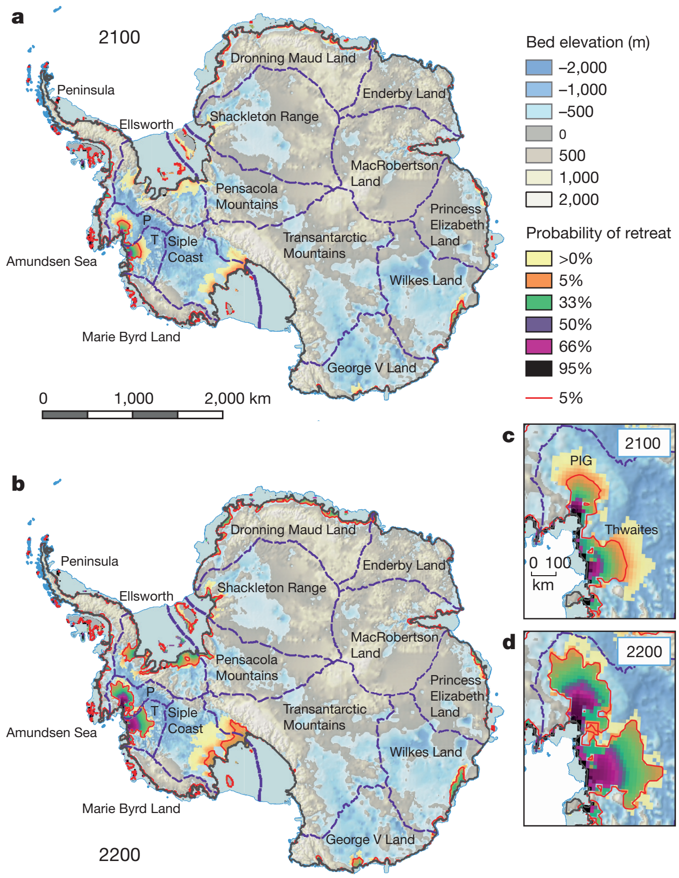


Figure 1 | Projected grounding-line retreat. a, **b**, Probability density estimates of grounding-line retreat at 2100 (**a**) and 2200 (**b**), overlaid on bedrock topography[24]. Red lines show 0.05 contour: an estimated 95% probability that retreat will be less extensive than this. **c**, **d**, ASE with Pine Island (PIG) and Thwaites glaciers.

combination of topography (downsloping bedrock below sea level) and low friction (Extended Data Fig. 2c–e). Our 95% quantiles for the ASE are 25 cm at 2100 and 48 cm at 2200 (all values are sea-level equivalent and, unless specified otherwise, 95% quantiles). The Thwaites region, which includes the Smith and Kohler glaciers[4], contributes the greater part of this: 58% at 2100 and 53% at 2200. This is partly due to the basin definition, but is also due to relatively rapid and substantial thinning of Thwaites upstream of the grounding line (see Supplementary Video 1). The Peninsula and Marie Byrd Land hardly respond, despite being assigned the same probabilities of onset as the ASE (owing to observed grounding-line retreat and thinning[4,16,17]), because their bedrock is largely above sea level.

Although basin contributions depend partly on coastline length, similar topographical limits are seen elsewhere: on the basis of projected ice-shelf surface and basal melting[7,18], Princess Elizabeth Land and MacRobertson Land are assigned substantial probabilities of MISI but contribute only 1 cm by 2200, while Dronning Maud Land is assigned lower probabilities but contributes up to 4 cm by 2100 and 8 cm by 2200. Responses also vary across the three basins of the Ronne–Filchner sector, which are assigned identical onset dates on the basis of projected Circumpolar Deep Water intrusion[6]. Ellsworth shows widespread ungrounding, with the 95% quantile at 2200 approximately delineating a previously deglaciated region[19] (Fig. 1 and Extended Data Fig. 3a), and contributes 9 cm by 2200; Shackleton Range and Pensacola Mountains show much less retreat and contribute 6 cm and 4 cm, respectively.

For Totten Glacier in Wilkes Land, our results suggest that if current dynamic thinning is MISI driven by Circumpolar Deep Water[8], the region has some potential for ungrounding (up to 5 cm by 2200). The Siple Coast is assigned a small probability from ice-shelf basal melting[18] but, when triggered, ungrounding is widespread owing to low basal friction (Extended Data Fig. 2c); we estimate that the total risk is small (up to 3 cm by 2200). These constraints are not absolute bounds—greater deglaciation has occurred in the past over longer time scales[9]—but appear to limit the amount of ice that can be lost in two centuries. Extended Data Figure 4 illustrates the effects of the two ice dynamical conditions, for example in George V Land, which is thought to be vulnerable in the long term[5] (Supplementary Information, section 2.2.1).

The total continental contribution to sea level is relatively low in the first century and accelerates in the second (Fig. 2a), although a second mode emerges at 6 cm by 2200 (Fig. 2b). The probability of exceeding 10 cm rises rapidly this century to 57% at 2100; for exceeding half a metre, it reaches only 33% at 2200 (Fig. 2c, d).

We find that the rate of sea-level rise from the ASE could be substantial this century: up to 1.3 mm yr^{-1} by 2050 and 2.1 mm yr^{-1} by 2100 (Fig. 3). However, many simulations stop (near zero mode at 2100 and local mode at 2200; Fig. 3b) or slow their retreat, particularly those with a linear–viscous friction law, so the 95% quantile at 2200 (1.1 mm yr^{-1}) is half that at 2100. Narrow zones of higher friction (hard bedrock) situated a few tens of kilometres upstream impede further retreat (Extended Data Fig. 3b). Extended Data Figure 5 shows this and other threshold behaviour dependent on friction law.

The strong dependence of ASE response on basal friction law lies behind the bimodal projections for Antarctica at 2200 (Extended Data Fig. 6). Projections of MISI using one friction law[2,3,10] may systematically under- or overestimate sea-level rise and will almost certainly underestimate its uncertainty. Although the sensitivity of

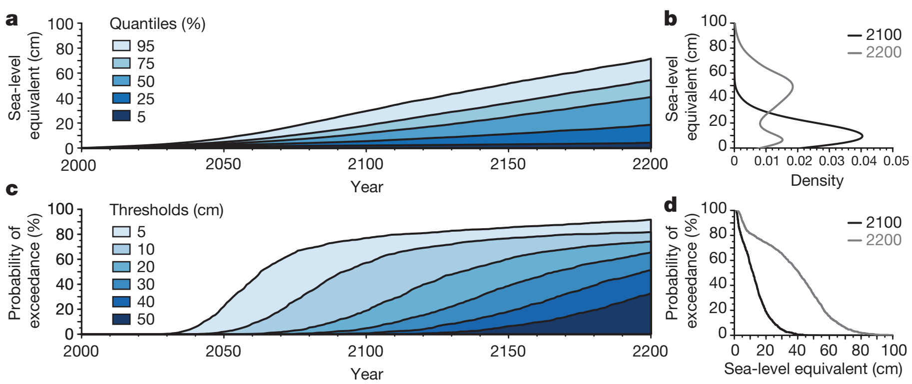


Figure 2 | Projected sea-level rise. a, Quantiles of Antarctic dynamic mass losses in cm sea-level equivalent as a function of time. **b**, Probability densities at 2100 and 2200. **c**, Probabilities of exceeding particular thresholds as a function of time. **d**, Probability of exceeding any threshold at 2100 and 2200.

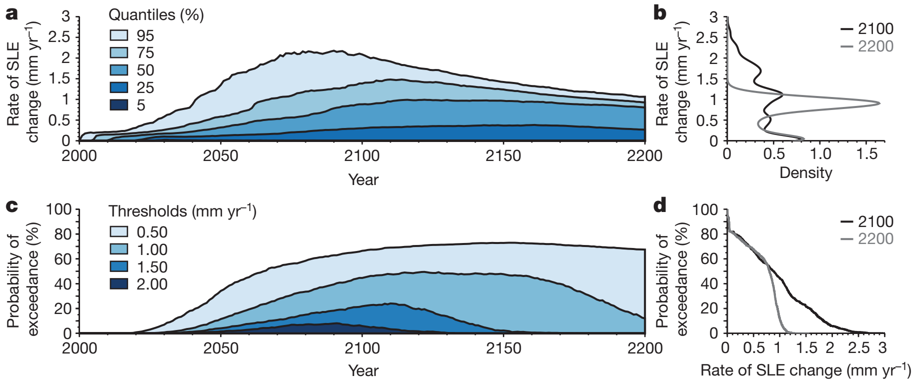


Figure 3 | Projected rate of sea-level rise from the Amundsen Sea. **a**, Quantiles of the rate of ASE dynamic mass losses in mm yr^{-1} sea-level equivalent (SLE) as a function of time. **b**, Probability densities at 2100 and 2200. **c**, Probabilities of exceeding particular thresholds as a function of time. **d**, Probability of exceeding any threshold at 2100 and 2200.

grounding-line migration to friction law has been explored previously[2,13,14], a fully Bayesian approach allows us to quantify the probabilistic contribution to uncertainty in sea-level rise. Extensive observations of basal type and hydrology, and better theoretical understanding of basal hydrology and sliding, would be needed to reduce this uncertainty.

Sensitivity to onset probabilities is limited for most basins by glaciological constraints that slow or stop retreat (Supplementary Information, section 2.2.2). Altering retreat onset probabilities by ±20% changes basin 95% quantiles at 2200 by up to about 1 cm, and using early or late ASE onset dates (2000–2010 or 2020–2030) changes the 95% quantile at 2200 by less than 2 cm (Extended Data Fig. 9a). Only Shackleton, Siple Coast and Transantarctic Mountains (Extended Data Fig. 9b–d) approach a linear response; increasing Siple Coast onset probabilities tenfold increases the 95% quantile at 2200 by 8 cm.

Observational calibration reduces projected quantiles by constraining the maximum rate of retreat and the regions over which this can occur (Extended Data Figs 7 and 8), mainly in the ASE. It presupposes that the best parameter values in one region are the best everywhere (although not the sliding law, which is not calibrated because it varies spatially; Supplementary Information, section 1.7). To assess the effect of this, we estimate that calibrating only the ASE contribution would increase 95% quantiles by approximately 6 cm (22%) at 2100 and 21 cm (29%) at 2200. Results are robust to other calibration choices (95% quantiles at 2200 vary by a few centimetres; Supplementary Information, section 2.2.4).

Our results are consistent with regional high-resolution model projections. In particular, projected ice losses by 2200 under A1B driven by one of the ocean simulations on which we base our onset probabilities[10] lie within our uncertainty estimates for the ASE (19–30% quantiles), Ronne–Filchner (Ellsworth, Pensacola Mountains, Shackleton: 56–65% quantiles) and Ross basins (Siple Coast, Transantarctic Mountains: 90%; tenfold Siple Coast probabilities 80%). For Marie Byrd Land, the high-resolution projections are lower than our ensemble, but the contribution to our result is less than a centimetre. Projected rates for Pine Island and Thwaites glaciers are also consistent with high-resolution modelling under idealized basal melting scenarios, and continental totals with a statistically based projection assuming ASE collapse in 2012 and linear growth of ice discharge elsewhere[20] (Supplementary Information, section 2.1).

Our projections are essentially incompatible with upper-bound estimates for MISI[9,21] of around 50–80 cm by 2100 and 140 cm by 2200 derived from physical arguments, extrapolation or low-resolution numerical models, and around 1 m by 2100 (95% quantile) from expert elicitation[22]. Half a metre of sea level rise by 2100 is not exceeded at the 99.9% quantile (uncalibrated: 98% percentile). Contributions of around 1 metre by 2100 were obtained (Extended Data Fig. 10 and Supplementary Information, section 2.2.3) by setting the parameter values to maximize ice loss and additionally either violating the theoretical limit or triggering immediate MISI everywhere (in 2000 for the Peninsula, ASE and Marie Byrd Land; 2020 elsewhere), but we do not consider these realistic. One metre by 2200 is exceeded at the 99.9% quantile (uncalibrated: 95% percentile).

We therefore find that MISI in the ASE could drive large and rapid sea-level rise but that the total Antarctic contribution is moderated by important physical constraints. Large uncertainties remain, in particular basal friction and its evolution, and further observations of surface and grounding-line changes would improve initialization and calibration. Future advances (high-resolution simulation of the ice-sheet–ice-shelf–ocean system; increased computational resources) will improve representation of the processes we parameterize and allow ensemble methods, while comparing multiple models would explore other representations of ice dynamics. But, given current understanding, our results indicate that plausible predictions of Antarctic ice-sheet instability leading to greater than around half a metre of sea-level rise by 2100 or twice that by 2200 would require new physical mechanisms[23], new projections of MISI triggers, or both.

Received 23 March; accepted 21 October 2015.
Published online 18 November 2015.

1. Schoof, C. Ice sheet grounding line dynamics: steady states, stability, and hysteresis. *J. Geophys. Res.* **112,** F03S28 (2007).
2. Favier, L. *et al.* Retreat of Pine Island Glacier controlled by marine ice-sheet instability. *Nature Clim. Chang.* **5,** 1–5 (2014).
3. Joughin, I., Smith, B. E. & Medley, B. Marine ice sheet collapse potentially under way for the Thwaites Glacier Basin, West Antarctica. *Science* **344,** 735–738 (2014).
4. Rignot, E., Mouginot, J., Morlighem, M., Seroussi, H. & Scheuchl, B. Widespread, rapid grounding line retreat of Pine Island, Thwaites, Smith, and Kohler glaciers, West Antarctica, from 1992 to 2011. *Geophys. Res. Lett.* **41,** 3502–3509 (2014).
5. Mengel, M. & Levermann, A. Ice plug prevents irreversible discharge from East Antarctica. *Nature Clim. Chang.* **4,** 451–455 (2014).
6. Hellmer, H. H., Kauker, F., Timmermann, R., Determann, J. & Rae, J. Twenty-first-century warming of a large Antarctic ice-shelf cavity by a redirected coastal current. *Nature* **485,** 225–228 (2012).
7. Kuipers Munneke, P., Ligtenberg, S. R. M., Van den Broeke, M. R. & Vaughan, D. G. Firn air depletion as a precursor of Antarctic ice-shelf collapse. *J. Glaciol.* **60,** 205–214 (2014).
8. Khazendar, A., Schodlok, M. P., Fenty, I., Ligtenberg, S. R. M., Rignot, E. & Van Den Broeke, M. R. Observed thinning of Totten Glacier is linked to coastal polynya variability. *Nature Commun.* **4,** 2857 (2013).
9. Intergovernmental Panel on Climate Change. *Working Group I Contribution to the IPCC Fifth Assessment Report Climate Change 2013: The Physical Science Basis* (Cambridge Univ. Press, 2013).
10. Cornford, S. L. *et al.* Century-scale simulations of the response of the West Antarctic Ice Sheet to a warming climate. *Cryosphere* **9,** 1579–1600 (2015).
11. Levermann, A. *et al.* Projecting Antarctic ice discharge using response functions from SeaRISE ice-sheet models. *Earth Syst. Dynam.* **5,** 271–293 (2014).

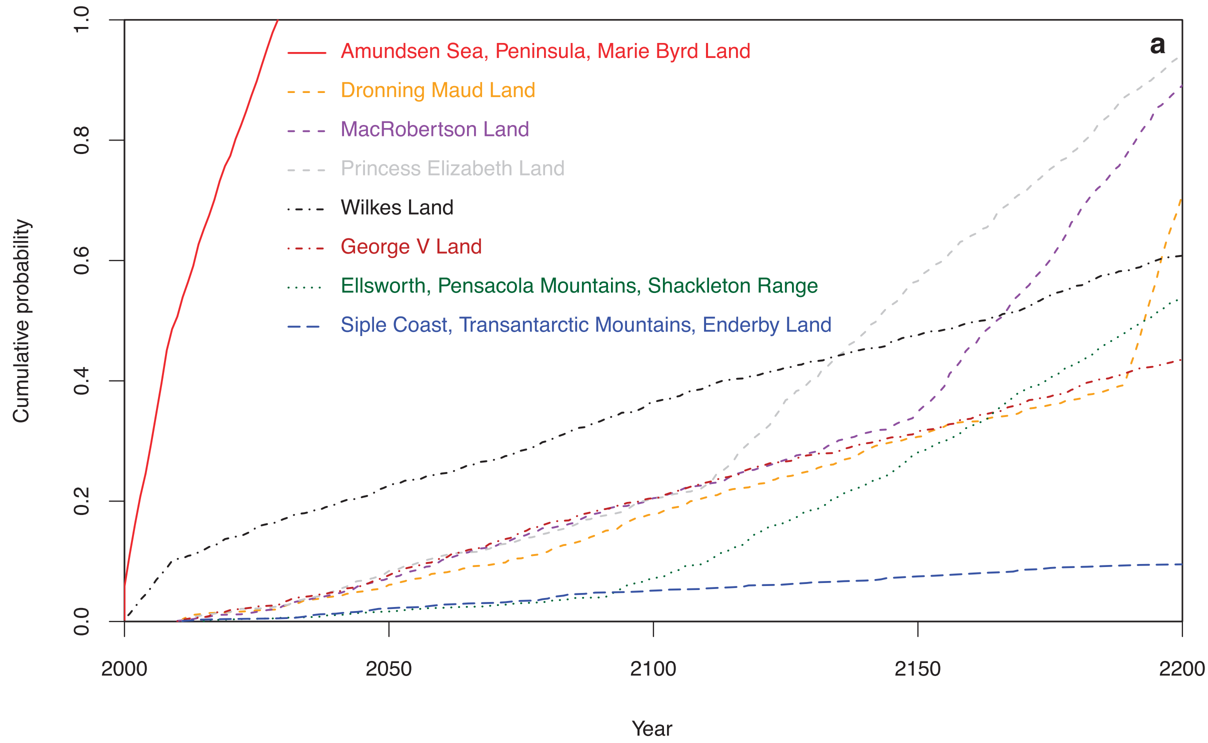
12. Gudmundsson, G. H., Krug, J., Durand, G., Favier, L. & Gagliardini, O. The stability of grounding lines on retrograde slopes. *Cryosphere* **6,** 1497–1505 (2012).
13. Joughin, I. *et al.* Basal conditions for Pine Island and Thwaites Glaciers, West Antarctica, determined using satellite and airborne data. *J. Glaciol.* **55,** 245–257 (2009).
14. Joughin, I., Smith, B. E. & Holland, D. M. Sensitivity of 21st century sea level to ocean-induced thinning of Pine Island Glacier, Antarctica. *Geophys. Res. Lett.* **37,** L20502 (2010).
15. Ritz, C., Rommeleare, V. & Dumas, C. Modeling the evolution of Antarctic ice sheet over the last 420,000 years: implications for altitude changes in the Vostok region. *J. Geophys. Res.* **106,** 31943–31964 (2001).
16. Pritchard, H. D., Arthern, R. J., Vaughan, D. G. & Edwards, L. A. Extensive dynamic thinning on the margins of the Greenland and Antarctic ice sheets. *Nature* **461,** 971–975 (2009).
17. Park, J. W. *et al.* Sustained retreat of the Pine Island Glacier. *Geophys. Res. Lett.* **40,** 2137–2142 (2013).
18. Timmermann, R. & Hellmer, H. H. Southern Ocean warming and increased ice shelf basal melting in the twenty-first and twenty-second centuries based on coupled ice-ocean finite-element modelling. *Ocean Dyn.* **63,** 1011–1026 (2013).
19. Ross, N. *et al.* Steep reverse bed slope at the grounding line of the Weddell Sea sector in West Antarctica. *Nature Geosci*. **5,** 393–396 (2012).
20. Little, C. M., Oppenheimer, M. & Urban, N. M. Upper bounds on twenty-first-century Antarctic ice loss assessed using a probabilistic framework. *Nature Clim. Chang.* **3,** 654–659 (2013).
21. Katsman, C. A. *et al.* Exploring high-end scenarios for local sea level rise to develop flood protection strategies for a low-lying delta—the Netherlands as an example. *Clim. Change* **109,** 617–645 (2011).
22. Bamber, J. L. & Aspinall, W. P. An expert judgement assessment of future sea level rise from the ice sheets. *Nature Clim. Chang*. **3,** 424–427 (2013).
23. Pollard, D., Deconto, R. M. & Alley, R. B. Potential Antarctic Ice Sheet retreat driven by hydrofracturing and ice cliff failure. *Earth Planet. Sci. Lett.* **412,** 112–121 (2015).
24. Le Brocq, A. M., Payne, A. J. & Vieli, A. An improved Antarctic dataset for high resolution numerical ice sheet models (ALBMAP v1). *Earth Syst. Sci. Data* **2,** 247–260 (2010).

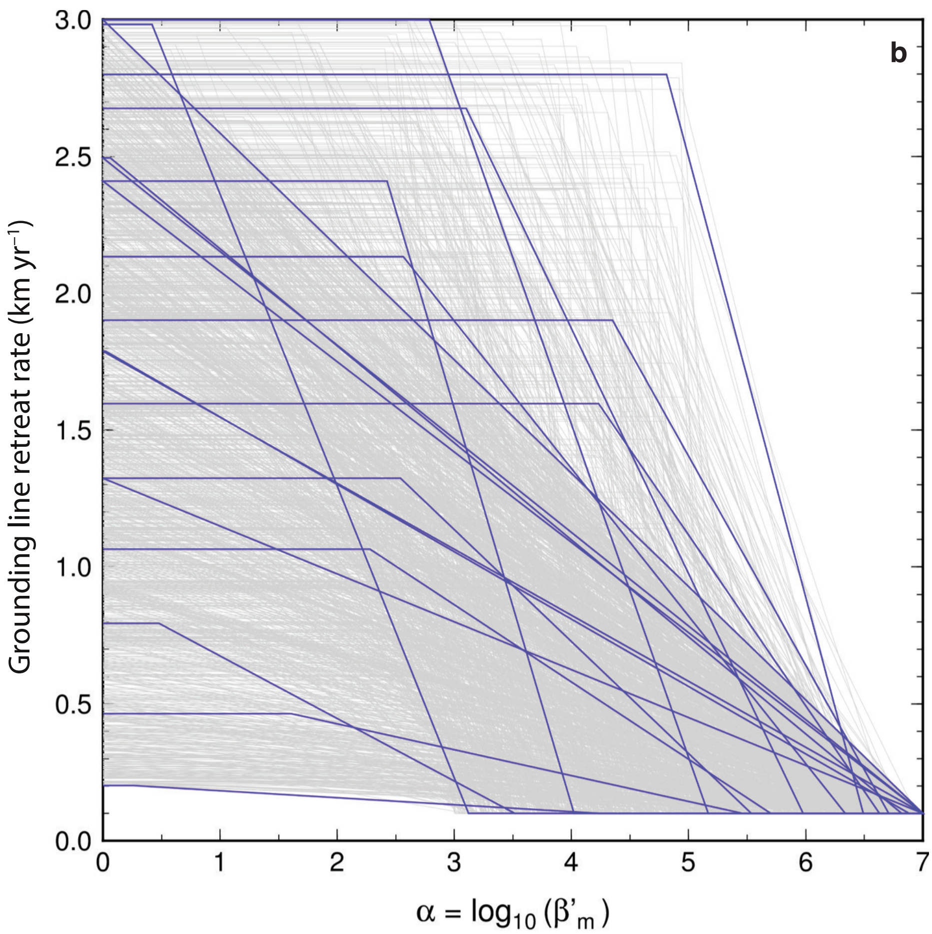
Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the ice2sea project funded by the European Commission's 7th Framework Programme through grant number 226375 (ice2sea contribution number ice2sea119), the UK National Centre for Earth Observation, NERC iGlass project, NERC and UK Met Office Joint Weather and Climate Research Programme, and the French National Research Agency (ANR) under the SUMER (Blanc SIMI 6) 2012 project ANR-12-BS06-0018. Most of the computations were performed using the CIMENT infrastructure (https://ciment.ujf-grenoble.fr), which is supported by the Rhône-Alpes region (grant CPER07 13 CIRA; http://www.ci-ra.org). We thank A. Shepherd and M. McMillan for observational data, H. Hellmer and R. Timmerman for model projection data, D. Vaughan and H. Hellmer for discussions about retreat onset, and J. C. Rougier for discussions about experimental design and calibration.

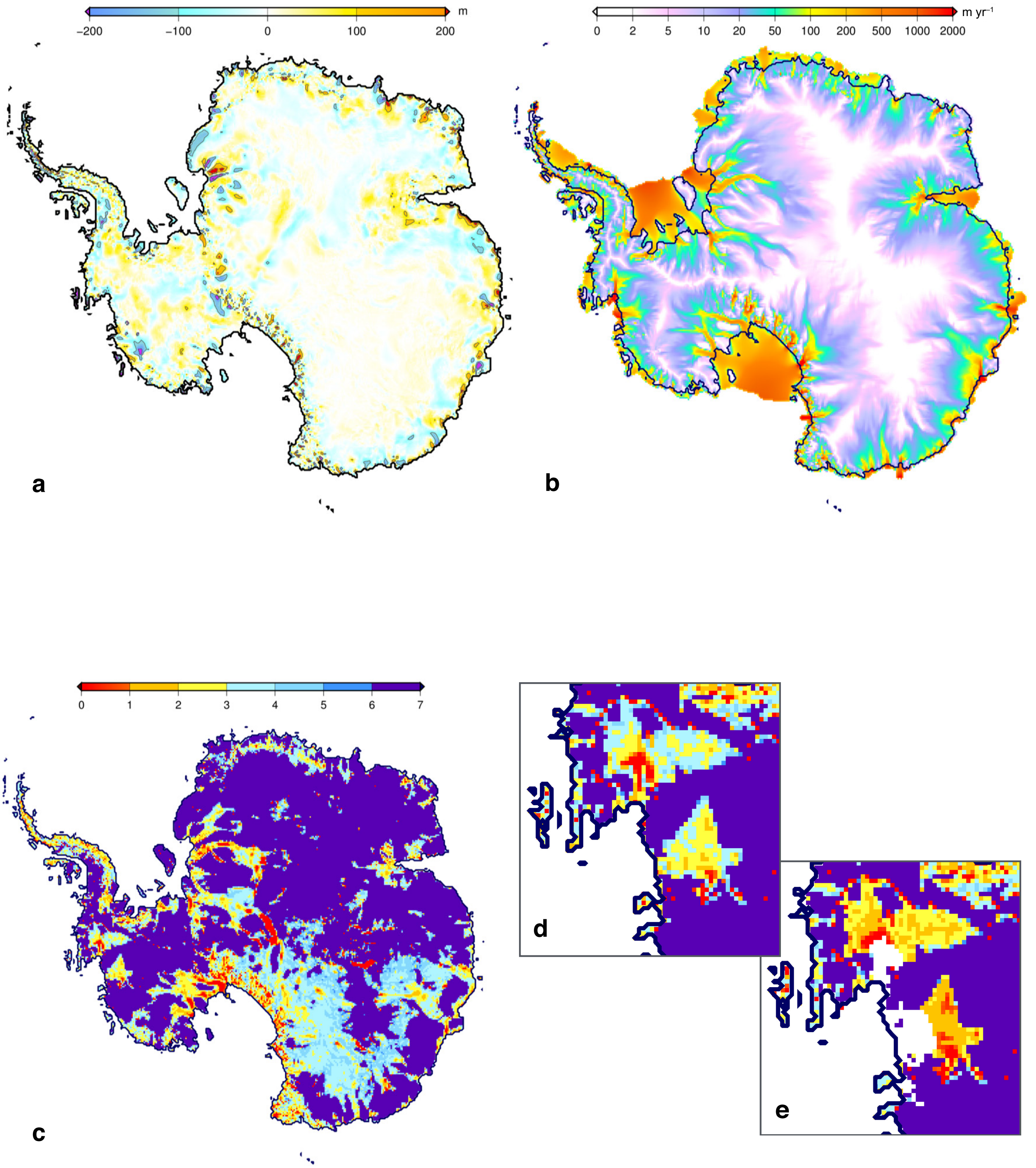
Author Contributions C.R. and V.P. worked on the development of the GRISLI model and did the numerical modelling. C.R. and G.D., with contributions from T.L.E., performed the physics analysis. T.L.E. designed the experiments with contributions from all authors, wrote the manuscript with contributions from C.R. and G.D., and performed the statistical analysis. T.L.E. and C.R. produced the figures and animation, with contributions from G.D. The sampling and geostatistical analysis were produced by A.J.P., and the theoretical conditions of grounding-line retreat were developed by R.A.H., C.R. and G.D.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.L.E. (tamsin.edwards@open.ac.uk).

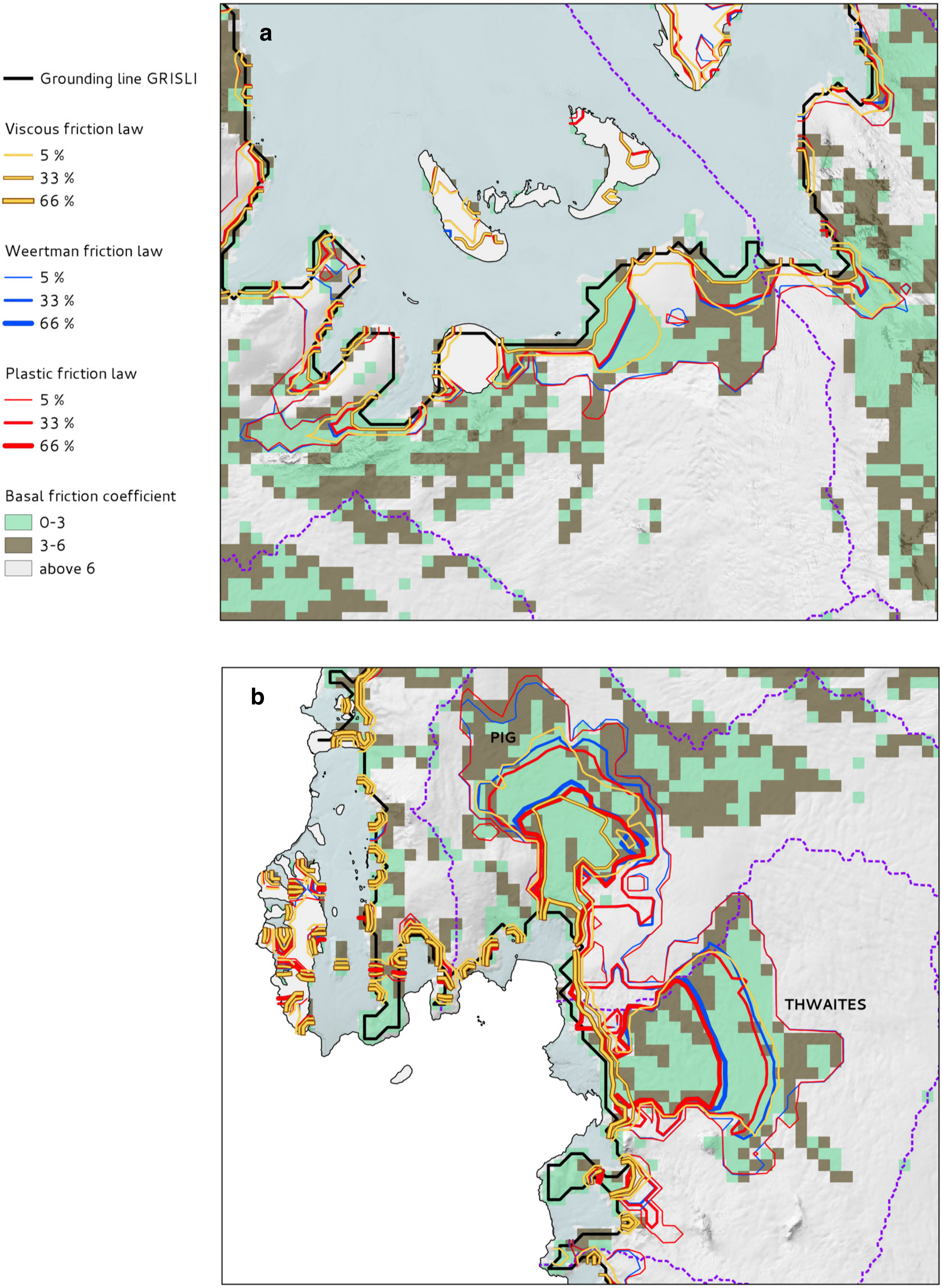




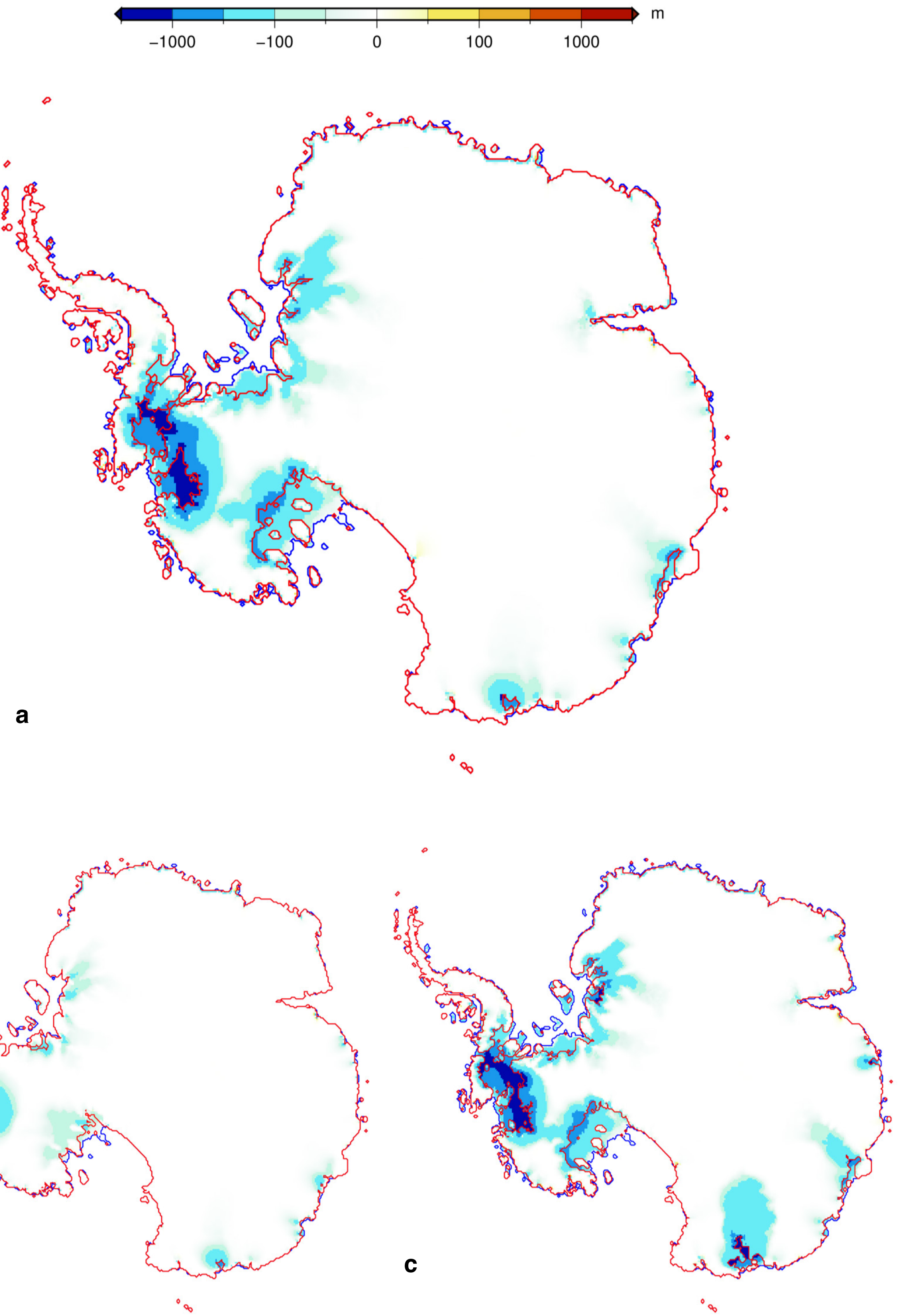
Extended Data Figure 1 | Grounding-line retreat parameterization. **a**, Cumulative probability distributions of MISI onset for 14 basins (Fig. 1) aggregated into 11 independent sectors. **b**, Piecewise linear parameterization prescribing the dependence of grounding-line retreat rate on the logarithm of the effective basal friction coefficient (Extended Data Fig. 2). Each of the 1,000 functional forms is a variant used in the ensemble; a subset are shown in bold as examples. See also Supplementary Information, sections 1.6.1, 1.6.2.



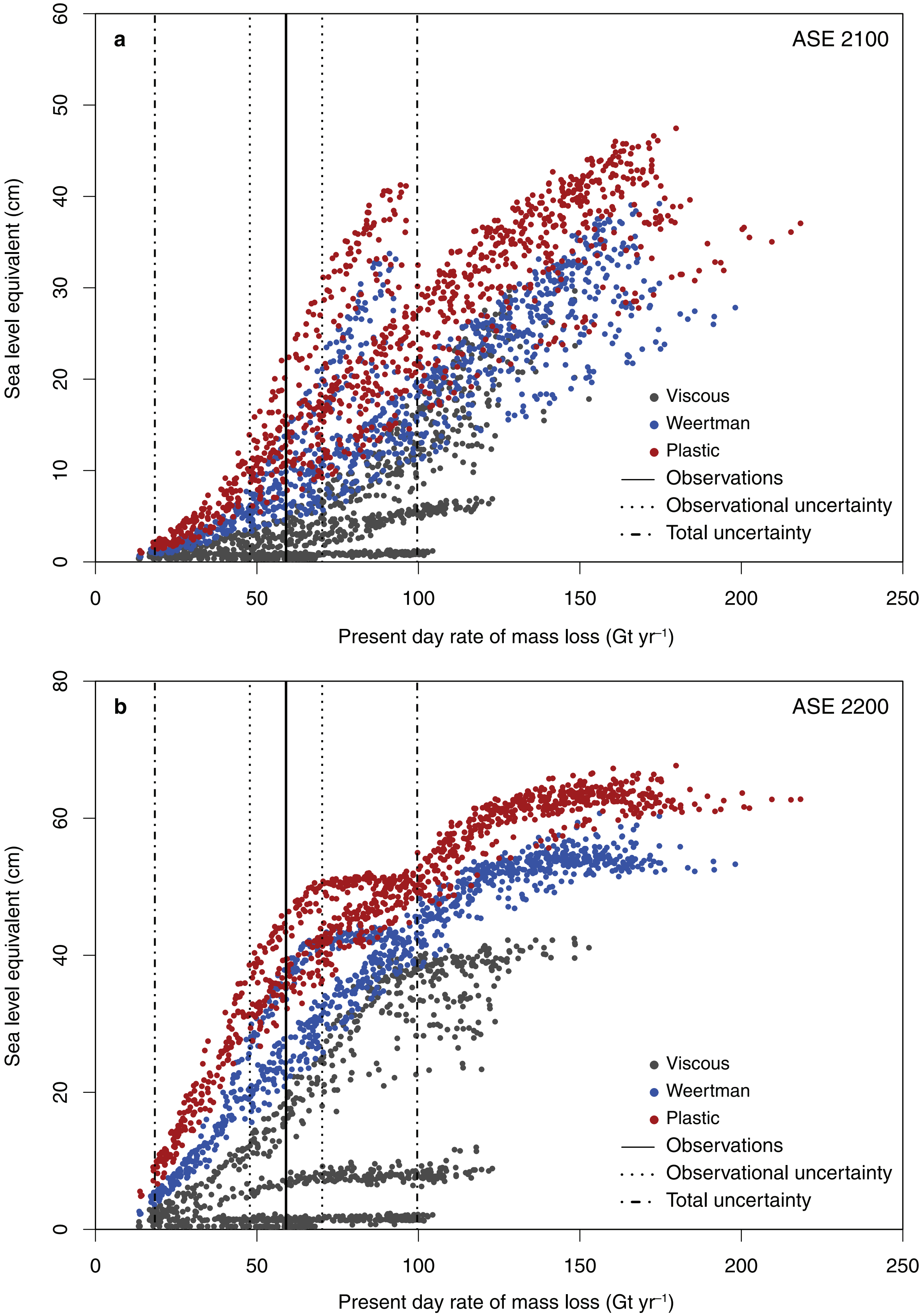
Extended Data Figure 2 | Initialization and basal friction evolution. **a**–**c**, Initial values of the difference between simulated and observed surface elevation (**a**); velocities averaged over ice thickness (**b**); the logarithm of the initial effective basal friction coefficient, $\alpha = \log_{10}(\beta_1'(\mathbf{x}; t = t_0))$ (**c**). **d**, As for **c**, showing the ASE. **e**, As for **d**, at 2200 in the plastic sliding law ensemble member that best matches present day ASE observations. See also Supplementary Information, section 1.5.



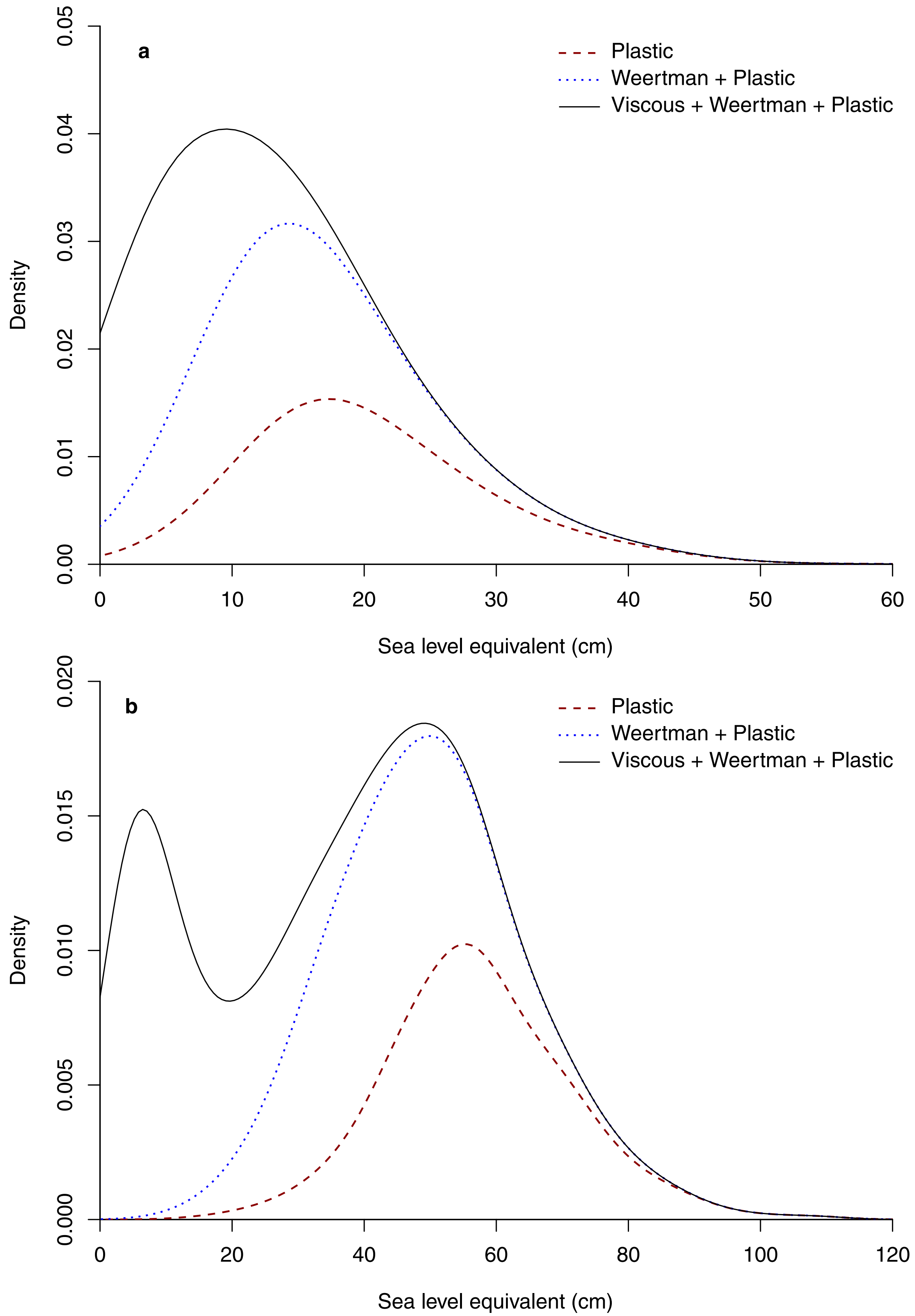
Extended Data Figure 3 | Projected grounding-line retreat and initial basal friction. a, b, Initial grounding line and map of α values (Extended Data Fig. 2) with retreat probability contours at 2200 for the Weddell Sea sector (**a**) and ASE (**b**): for example, there is an estimated 33% probability that grounding-line retreat will be less extensive than the 66% contour.



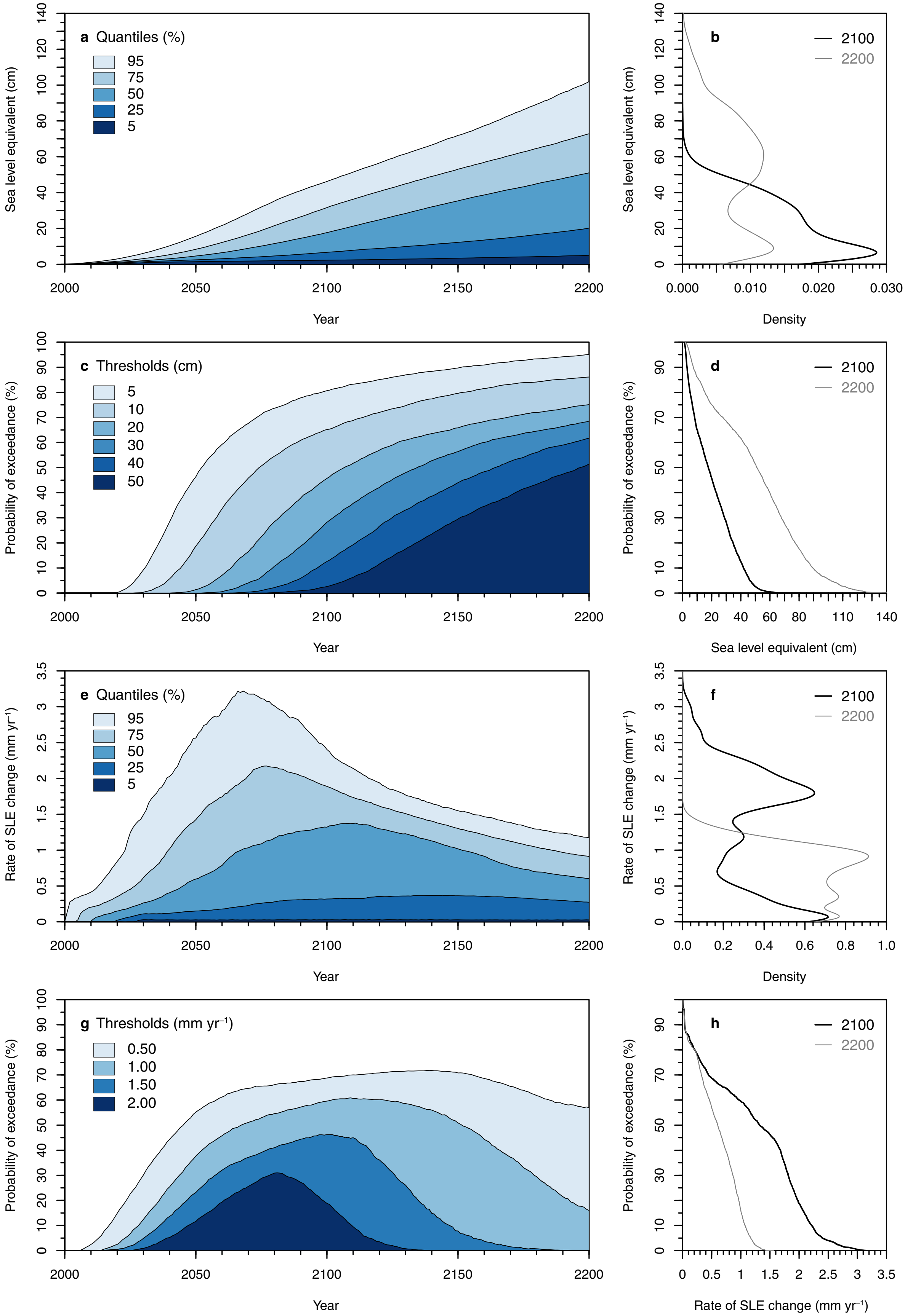
Extended Data Figure 4 | Ice dynamical conditions for retreat. **a**–**c**, Surface elevation changes at 2200 in the ensemble member with maximum sea-level contribution at 2200 (plastic sliding law): standard settings (**a**); 'Schoof flux' condition off, thereby only allowing grounding-line retreat along strictly downsloping bedrock (**b**); 'no suction' check off, thereby allowing thinning due to grounding line retreat to occur faster than the theoretical limit (**c**). See also Supplementary Information, section 2.2.1.



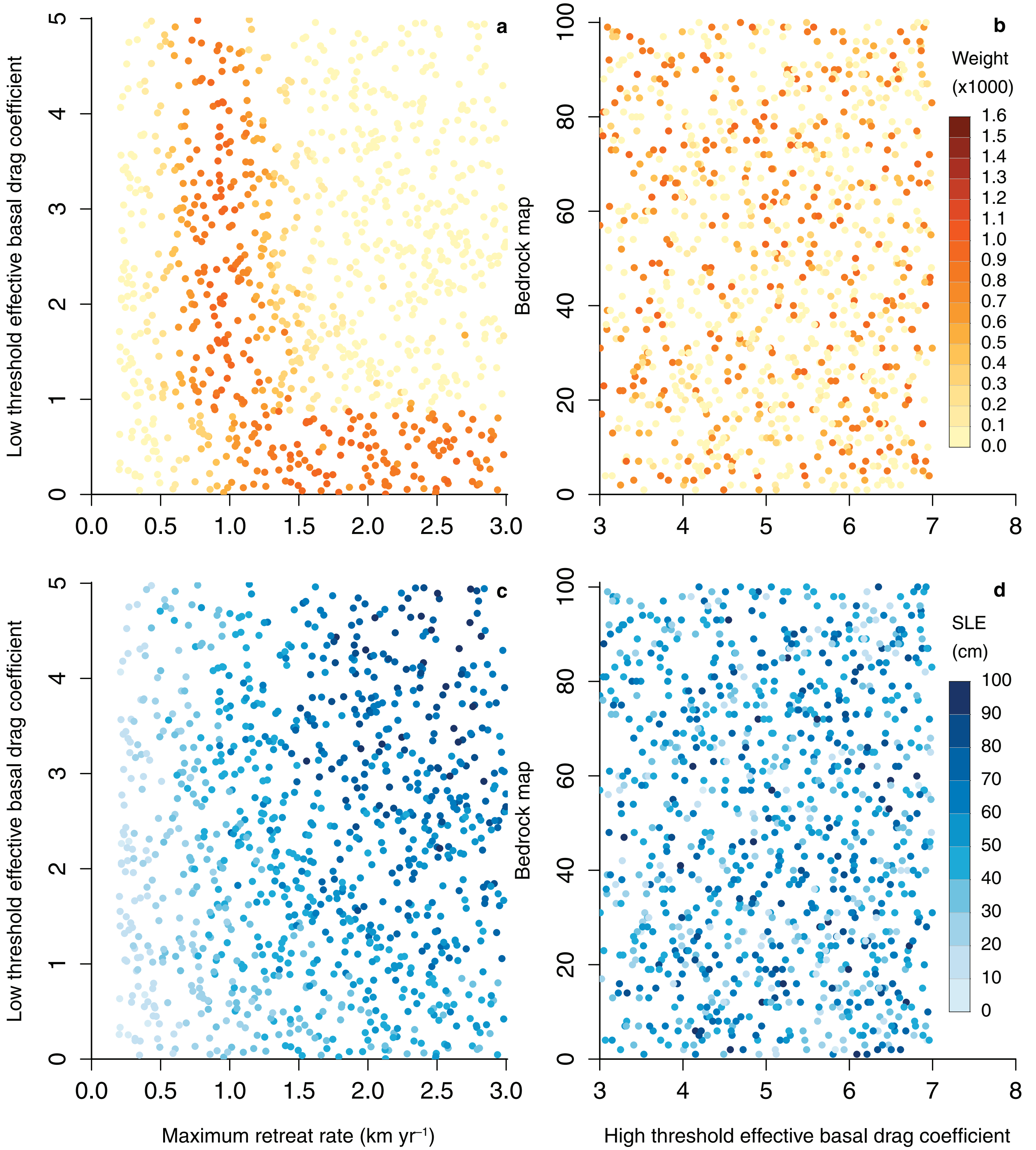
Extended Data Figure 5 | Relationship between present and future sea-level contributions from the Amundsen Sea. a, b, Dynamic mass losses in cm sea-level equivalent from the ASE at 2100 (**a**) and 2200 (**b**), as a function of present day mass loss in the same region. The branches arise from interactions between basal drag coefficient and friction law that produce different rates of, and impediments to, grounding-line retreat. The observed mass loss is shown, along with observational ($\pm 3\sigma_o$) and total ($\pm 3\sigma_t$) uncertainties (Supplementary Information, section 1.7).



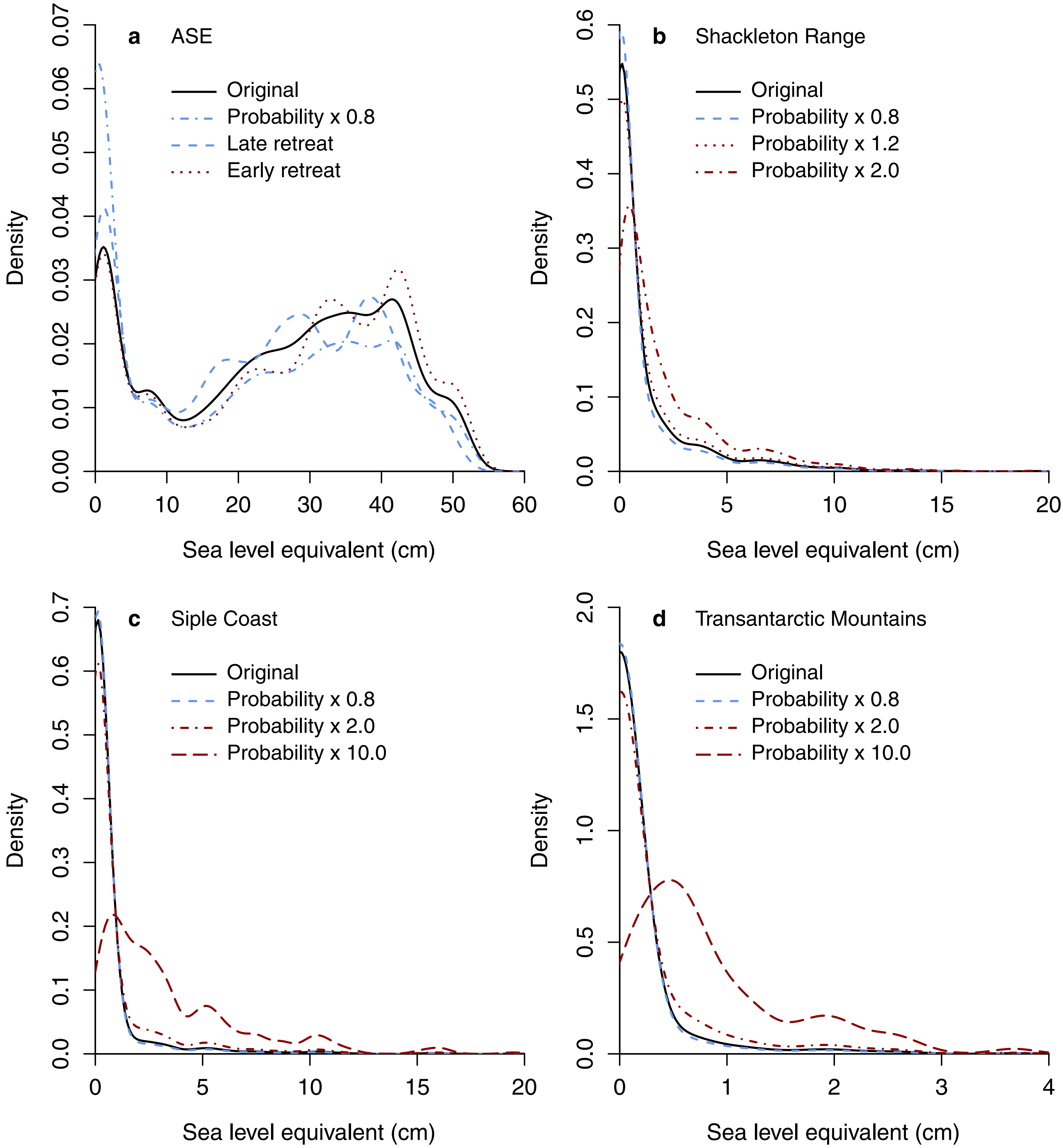
Extended Data Figure 6 | Contributions of each basal friction law. a, b, Probability distributions of Antarctic dynamic mass losses in cm sea-level equivalent at 2100 (**a**) and 2200 (**b**) (as in Fig. 2b), showing the cumulative contributions of the basal friction laws.



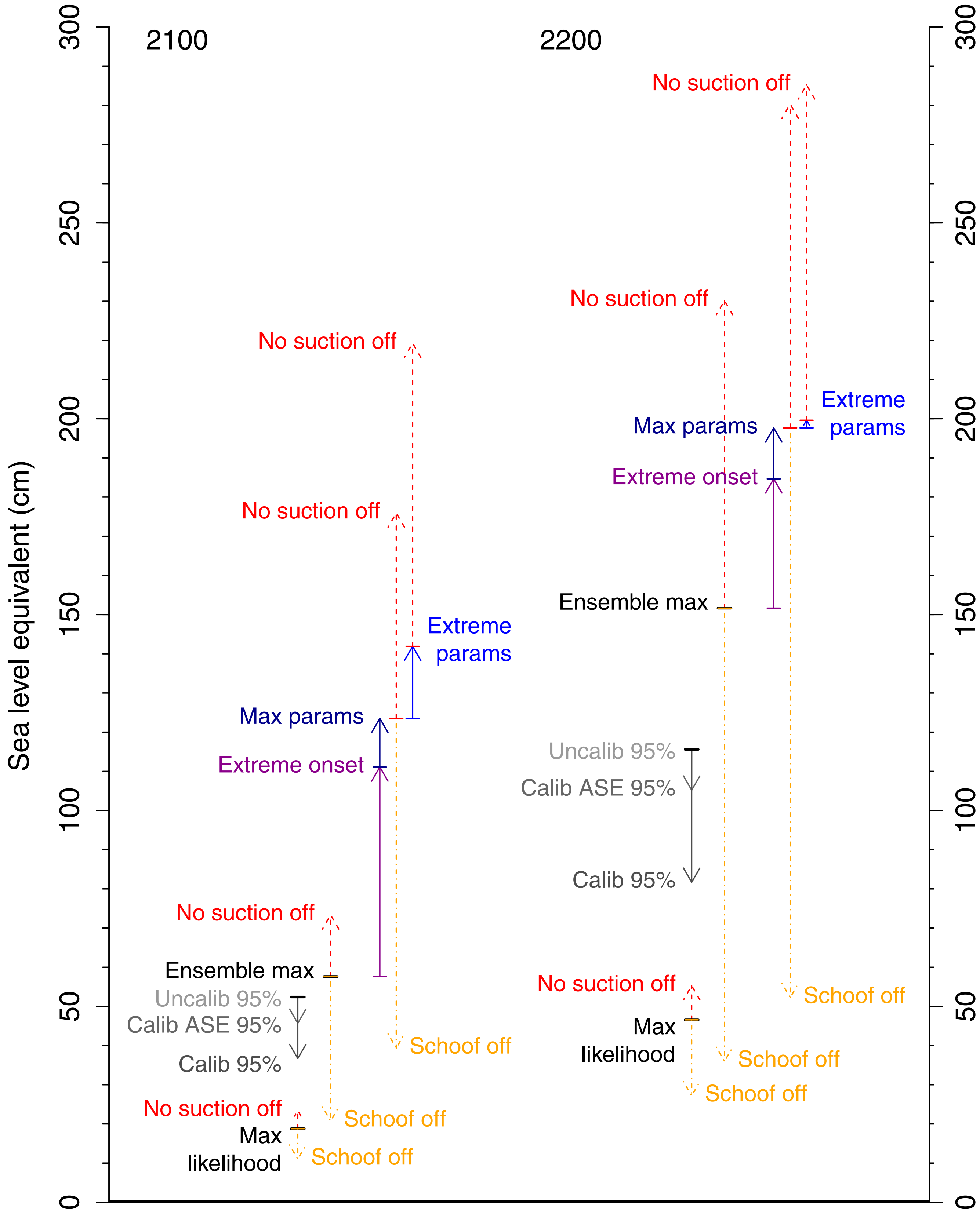
Extended Data Figure 7 | Uncalibrated projections. a–h, Prior (uncalibrated) projections of Antarctic dynamic mass losses in cm sea-level equivalent (**a**–**d**); rate of ASE dynamic mass losses in mm yr^{-1} sea-level equivalent (SLE) (**e**–**h**). Posterior (calibrated) projections are in Figs 2 and 3. See also Supplementary Information, section 1.7.



Extended Data Figure 8 | Parameter calibration and influence. **a**, **b**, Weights for each of the 1,000 sub-ensemble parameter sets (averaged over basal friction laws) as a function of low threshold of effective basal drag coefficient (α_{low}) and maximum retreat rate (v_{max}) (**a**); bedrock map index and high threshold of effective basal drag coefficient (α_{high}) (**b**). Darker colours indicate values favoured by observational calibration. **c**, **d**, Uncalibrated dynamic mass losses at 2200 in cm sea-level equivalent (SLE) as functions of the same.



Extended Data Figure 9 | Sensitivity to retreat onset distributions. **a–d**, Projections at 2200 estimated for four individual basins under different retreat onset scenarios. **a**, ASE: original; 'probability × 0.8', in which 20% of simulations are set to zero contribution; 'late retreat', in which all simulations begin retreating between 2020 and 2030; and 'early retreat', retreating between 2000 and 2010; **b**, Shackleton; **c**, Siple Coast; and **d**, Transantarctic Mountains: original, and onset probabilities adjusted by the factors shown. See also Supplementary Information, section 2.2.2.



Extended Data Figure 10 | Sensitivity tests for plastic sliding law. Antarctic dynamic mass losses in cm sea-level equivalent under various conditions: 'Max likelihood', the plastic simulation that best matches present day ASE observations; 'Uncalib 95%' and 'Calib 95%', the plastic quantiles before and after calibration, respectively, and 'Calib ASE 95%', the estimate calibrating only the ASE; 'Ensemble max', the simulation with highest contribution at 2200; 'Extreme onset', the previous with all basins retreating from 2000 or 2020; 'Max params', the previous with retreat parameters at maximum values and 'Extreme params' at higher values; 'Schoof flux' and 'No suction' checks off (dashed to indicate that they are physically unrealistic). See also Supplementary Information, sections 2.2, 2.3.

doi:10.1038/nature15539

Death from drought in tropical forests is triggered by hydraulics not carbon starvation

L. Rowland[1], A. C. L. da Costa[2], D. R. Galbraith[3], R. S. Oliveira[4], O. J. Binks[1], A. A. R. Oliveira[2], A. M. Pullen[5], C. E. Doughty[6], D. B. Metcalfe[7], S. S. Vasconcelos[8], L. V. Ferreira[9], Y. Malhi[6], J. Grace[1], M. Mencuccini[1,10] & P. Meir[1,11]

Drought threatens tropical rainforests over seasonal to decadal timescales[1–4], but the drivers of tree mortality following drought remain poorly understood[5,6]. It has been suggested that reduced availability of non-structural carbohydrates (NSC) critically increases mortality risk through insufficient carbon supply to metabolism ('carbon starvation')[7,8]. However, little is known about how NSC stores are affected by drought, especially over the long term, and whether they are more important than hydraulic processes in determining drought-induced mortality. Using data from the world's longest-running experimental drought study in tropical rainforest (in the Brazilian Amazon), we test whether carbon starvation or deterioration of the water-conducting pathways from soil to leaf trigger tree mortality. Biomass loss from mortality in the experimentally droughted forest increased substantially after >10 years of reduced soil moisture availability. The mortality signal was dominated by the death of large trees, which were at a much greater risk of hydraulic deterioration than smaller trees. However, we find no evidence that the droughted trees suffered carbon starvation, as their NSC concentrations were similar to those of non-droughted trees, and growth rates did not decline in either living or dying trees. Our results indicate that hydraulics, rather than carbon starvation, triggers tree death from drought in tropical rainforest.

Drought-response observations from both field-scale experiments and natural droughts have demonstrated increased mortality over the short-term (1–3 years), with notably higher vulnerability for some taxa, and for larger trees[6,9,10]. After several years of drought, recovering growth rates in smaller trees, dbh (diameter at breast height) <40 cm, and reduced mortality have been recorded at different locations[6,11,12]. However, the long-term (>10 year) sensitivity of tropical forests to predicted prolonged and repeated water deficit[1–3] and the physiological mechanisms influencing this are poorly understood. Through-fall exclusion (TFE) studies, that create soil moisture deficit by the exclusion of a fraction of incoming rainfall, provide the only current means to assess the long-term response in mechanistic detail[5,13].

Trees experiencing drought stress are thought to die from direct physiological failure and/or from injury and biotic attack associated with a decline in physiological vigour[14]. A global effort to identify the relevant physiological mechanisms triggering death and thus to improve predictions of forest tree mortality has focused on the twin possibilities of: (1) failure to supply sufficient carbon substrate to metabolism following drought-related reductions in photosynthesis and increased use of NSC, theoretically leading to carbon starvation; and (2) deterioration of the water-conducting xylem tissue, causing a rapid or gradual failure of key dependent processes (for example, gas exchange, photosynthesis, phloem transport), and potentially leading to tissue desiccation[14,15], ultimately leading to mortality. Despite recent intensive research, it is unclear how important these two mechanisms are in different biomes and how, or whether, to model them[16].

Since 2002 a 50% TFE treatment has been implemented at a 1 ha-scale drought experiment in old-growth forest at Caxiuanã National Forest Reserve, Pará State, Brazil[6,12], to simulate maximum possible rainfall reductions predicted to occur in parts of Amazonia by 2100 (ref. 1). Mortality surveys, recruitment and growth rates of all trees ≥ 10 cm dbh, have been monitored through the experimental period (see Methods). Recently, seasonal data on NSC concentrations were measured on leaves, branches and stems of 41 trees (20 trees on the control, 21 trees on the TFE) of the most common genera in the experiment (Extended Data Table 1). Xylem vulnerability curves were also performed on the branches of these trees (see Methods). Here, we synthesize these data to test whether long-term soil moisture deficit alters NSC storage and use in tropical rainforest trees, and if this, or hydraulic processes, are most strongly associated with increased mortality rates.

By 2014, following 13 years of the TFE treatment, cumulative biomass loss through mortality was $41.0 \pm 2.7\%$ relative to pre-treatment values (Fig. 1a), and the rate of loss had increased substantially since the previous reported value of $17.2 \pm 0.8\%$, after 7 years of TFE[6]. Accelerating biomass loss and failure to recover substantially, or to reach a new equilibrium[13], has led to a committed flux to the atmosphere from decomposing necromass of 101.9 ± 19.1 Mg C ha^{-1} (Fig. 1a). This biomass loss has been driven by elevated mortality in the largest trees (Fig. 1b), as previously observed over shorter timescales[6], and has created a canopy that has had a persistently lower average leaf area index during 2010–2014 ($12.0 \pm 1.2\%$ lower; Extended Data Fig. 1).

Remarkably, individual tree growth rates for the four years before death showed no significant reduction in either the TFE or control plots (Fig. 2a), indicating that growth is prioritised to the point of death irrespective of the soil moisture deficit treatment. From 2008, tree growth in every wet season (January–June) on the TFE treatment relative to the control was significantly elevated ($P < 0.05$) in the small and medium trees (up to 4.6 ± 0.2 times higher in small trees, and 2.9 ± 0.2 times higher in medium trees), and maintained in the largest trees (10–20 cm, 20–40 cm and >40 cm dbh, respectively; Fig. 2b–d). Elevated wet season growth occurred despite 0.1–0.9 MPa reduction in average soil water potential (Ψ_s) at depths of 0–4 m on the TFE and a loss of seasonality in Ψ_s (Extended Data Fig. 2). Increased growth in the small trees occurred from 2008 onwards, following earlier substantial mortality of large trees (Fig. 1), which generated canopy gaps. Increased light availability to smaller trees and, presumably, reduced below-ground competition for water and nutrients, allowed competitive release of trees on the TFE[8], and elevated growth rates. Competitive release on the TFE implies that, following 13 years of drought-stress, photosynthetic production is sufficient not only to maintain growth in the largest trees (Fig. 2d), but to increase growth

[1]School of GeoSciences, University of Edinburgh, Edinburgh EH9 3FF, UK. [2]Centro de Geosciências, Universidade Federal do Pará, Belém 66075-110, Brazil. [3]School of Geography, University of Leeds, Leeds LS2 9JT, UK. [4]Instituto de Biologia, UNICAMP, Campinas 13.083-970, Brazil. [5]The University of Cambridge, Cambridge CB2 1TN, UK. [6]Environmental Change Institute, The University of Oxford, Oxford OX1 3QY, UK. [7]Department of Physical Geography and Ecosystem Science, Lund University, Lund S-223 62, Sweden. [8]EMBRAPA Amazônia Oriental, Belém 66095-903, Brazil. [9]Museu Paraense Emílio Goeldi, Belém 66077-830, Brazil. [10]ICREA at CREAF, 08193 Cerdanyola del Vallés, Spain. [11]Research School of Biology, Australian National University, Canberra, Australian Capital Territory 2601, Australia.

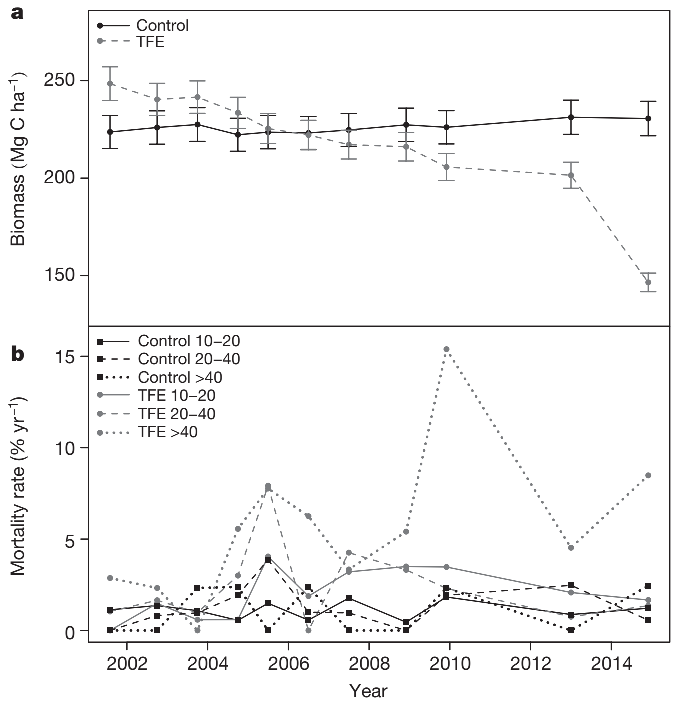


Figure 1 | Changes in biomass and mortality rates. a, Biomass on the control and TFE plot from 2001–2014 (Mg C ha^{-1} yr^{-1}). Error bars show the s.e.m. calculated from 12 estimates of biomass for trees on the control plot ($n = 369$) and TFE ($n = 358$), accounting for uncertainty in wood density and allometric equations (see Methods). **b**, Mortality rate (% stems per year) for trees on the control plot (black) and TFE (grey) separated for trees of 10–20 cm dbh (control $n = 164$–193, TFE $n = 132$–174, with range showing 2001–2014 maximum and minimum n), 20–40 cm (control $n = 97$–105, TFE $n = 81$–104), dbh and >40 cm (control $n = 41$–45, TFE $n = 17$–37). The genus and date of death for each tree used in the mortality rate calculations is shown in Extended Data Table 3.

in trees <40 cm dbh (Fig. 2b, c). This response would not be possible if the majority of trees were severely carbon limited, unless very considerable long-term (or renewed) carbon resources were being drawn upon.

Prioritization of growth under drought in the TFE is consistent with recent observations following short-term drought in Amazonia[7]. However, the maintenance of NSC concentrations in the TFE treatment suggests that the prioritisation of growth during drought does not occur at the expense of depleted carbon stores, as previously hypothesized[7]. Neither the concentrations of soluble sugar (carbon immediately available to metabolism) nor starch (stored carbon which can be converted to sugars) were significantly depleted in stem, leaf and branch tissue from the TFE, relative to control (Fig. 3). The seasonal changes in both sugar and starch concentrations, which varied by 50–90%, were much larger than any differences associated with the TFE treatment (Fig. 3). Despite 13 years of severely reduced soil moisture availability, the seasonal cycle and use of NSCs was unaltered, implying that the sampled trees did not draw significantly upon their NSC reserves to buffer against the long-term effects of soil moisture deficit. Large changes in carbon allocation from roots and leaves to maintain stem growth during drought[17] have not been reported on the TFE[12]. Similarly, no drought-induced reductions in photosynthetic capacity occurred on the TFE[18], although how total canopy productivity is affected remains uncertain. Considering this and additional evidence of no increase in herbivore attack on the TFE (Extended Data Fig. 3), our results suggest progressive carbon starvation and biotic foliar consumption are not important drivers of the mortality patterns observed in the TFE forest following extended severe soil moisture deficit (>10 years).

Deterioration of the water transport system in the xylem tissues following drought can also lead to death[17,19]. The vulnerability of the xylem to drought is described by a vulnerability curve[20], which relates water potential in xylem conduits to loss of hydraulic conductivity because of occlusions by gas emboli. The water potential at which 50% loss of xylem conductivity occurs (P_{50}, MPa) is a commonly used index of embolism resistance[20]. We determined xylem P_{50} for the trees on the control and TFE plots, with tree dbh ranging from 15 to 48 cm. A highly significant decrease in P_{50} with dbh was found across TFE and control (Extended Data Table 2, $P < 0.01$). As dbh increased from 15 to 48 cm there was a 1.3 ± 0.2 MPa reduction in the P_{50} value, with significant genus-to-genus differences (Fig. 4). Leaf water potential (Ψ_l) could only be measured during limited sampling campaigns (2–3 days) that were characterized by low vapour pressure deficit (VPD, 54–59% of peak dry season values) and unseasonal rainfall in the preceding days. Differences between treatment and control Ψ_l were not detected. Mean midday Ψ_l recorded across

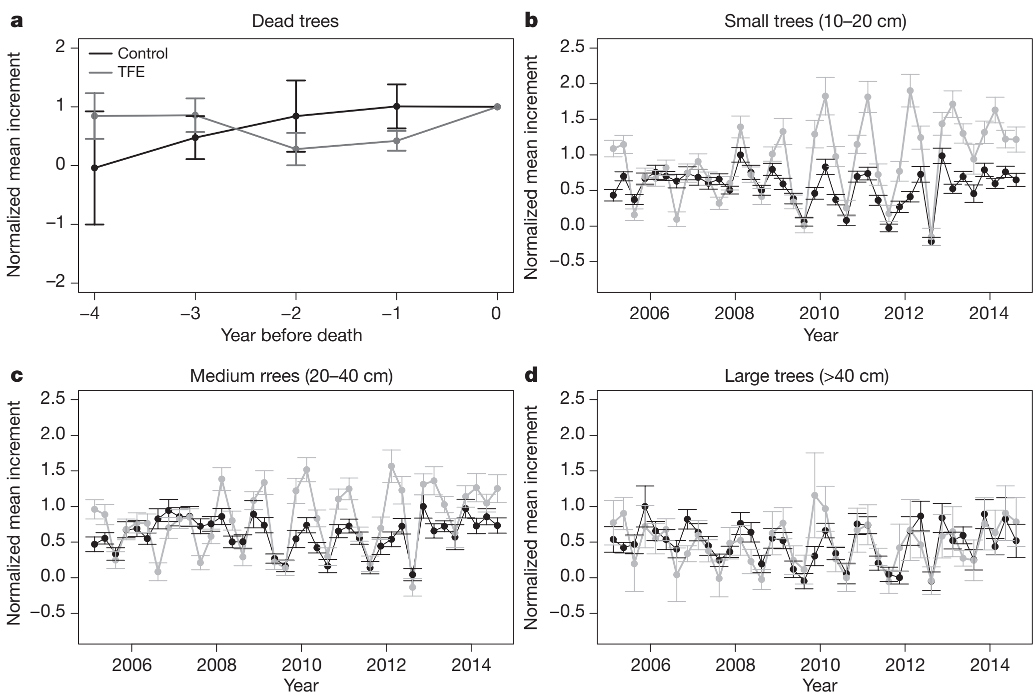


Figure 2 | Tree growth of living and dead trees. a, Average annual stem increment in the 4 years before death for all trees which died on the control ($n = 35$) and TFE ($n = 60$) from 2005 onwards. Growth rates are normalized by the growth in year 0, the year in which the tree died. **b**–**d**, Stem increment for small (**a**, 10–20 cm dbh, $n = 153$, $n = 116$), medium (**b**, 20–40 cm dbh, $n = 73$, $n = 78$) and large (**c**, >40 cm dbh, $n = 18$, $n = 8$) trees on the control and TFE plot respectively from 2005 to 2014. Values for **b**–**d** are normalized by the maximum increment on the control plot. Error bars show the s.e.m.

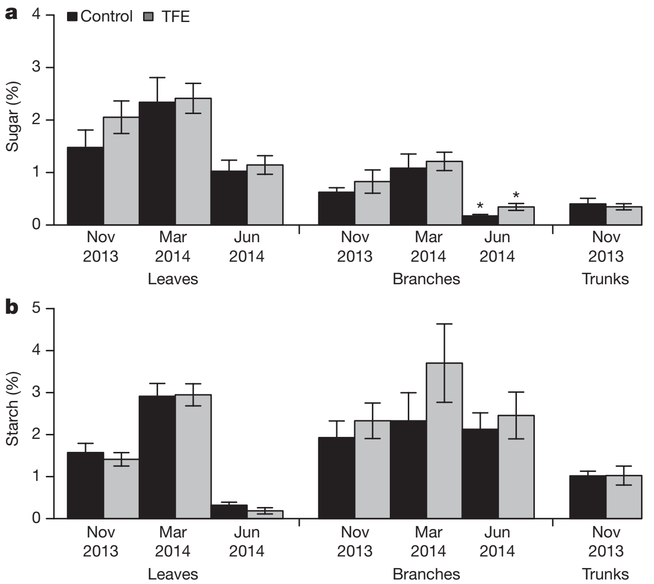


Figure 3 | Leaf, branch and stem NSC concentrations. a, b, Percentage of soluble sugars (**a**) and starch (**b**) in biomass of leaves, branches and stems, in the late dry season (November 2013), mid wet season (March 2014) and in the wet-to-dry transition (June 2014). Each value for leaves, branches or stems represents an average of samples taken from $n = 20$ trees on the control, and $n = 21$ trees on the TFE. TFE Error bars show s.e.m. and * indicates a significant difference at $P < 0.05$ using the Wilcoxon test. June 2014 has significantly elevated sugar values on the TFE plot; however, the absolute values for sugar concentration are very low, and the absolute differences are very small.

all the trees together with the vulnerability curves determined for each genus were used to predict the percentage loss of xylem conductivity (PLC) with dbh. Values of PLC at mean Ψ_l increased with dbh, with the largest diameter trees predicted to have reductions in conductive capacity of about 80% in some genera, indicating significant vulnerability to hydraulic deterioration (inset of Fig. 4).

Given no evidence of carbon starvation and similar Ψ_l across plots in the sample dates, why did many more trees die in TFE than control? The lack of treatment differences in Ψ_l contrasts starkly with the long-term records of lower Ψ_s (Extended Data Fig. 2). The lack of difference in midday Ψ_l could have been caused by sampling constraints or by isohydric behaviour. We found evidence of non-isohydric behaviour in our diurnal Ψ_l measurements (Extended Data Fig. 4), with overall strong linear declines in Ψ_l observed with increasing VPD on the control ($R^2 = 0.18$, $P < 0.01$) and in particular on the TFE ($R^2 = 0.33$, $P < 0.01$). Consequently, limited sampling is the most likely cause of equal Ψ_l between the two plots, with TFE trees likely to be having more negative Ψ_l and lower hydraulic conductance during VPD maxima in the dry season. Reduced carbon uptake because of stomatal closure in some TFE trees is possible[21], but is unlikely to have caused carbon starvation considering that growth rates were maintained or elevated on the TFE (Fig. 2) and that radial growth should decline before photosynthesis in drought conditions[22]. Even with an isohydric response, trees on the TFE would still be likely to suffer greater hydraulic deterioration caused by greater PLC in the roots and main stem. Strongly reduced Ψ_s on the TFE (Extended Data Fig. 2) and significant hydraulic vulnerability of the tall trees are consistent with the hypothesis of hydraulic deterioration as the most likely trigger of greater mortality, particularly in the largest trees, as observed. Why the xylem tissue of larger trees is more vulnerable to embolism deserves further study. Taller trees are predisposed to greater hydraulic stress, from elevated atmospheric demand and longer hydraulic path lengths[23]. As the canopies are exposed to rainfall in the TFE, smaller trees could avoid hydraulic deterioration through leaf water uptake[24–26], but this may not be sufficient to save the largest trees, which we hypothesize are forced to maintain their high growth rates until death to continually replace dysfunctional xylem.

Following decadal-scale soil moisture depletion, our results suggest that tropical rainforests will experience accelerating biomass loss and a likely transition to a lower statured, lower biomass forest state, due to substantially elevated mortality of the largest trees. This mortality is most likely triggered by hydraulic processes, which lead to hydraulic deterioration and subsequent, potentially rapid, limitations in carbon uptake[21], instead of being caused directly by gradual carbon starvation. Under natural drought these forests may be under greater risk than from experimental drought, as severe soil moisture deficit is combined with low humidity and high air temperature, increasing hydraulic demand. Improved prediction of the sensitivity of tropical tree mortality to drought should therefore focus on improved model simulation of plant hydraulics and modelling environmental controls on growth[22,27]. Decadal-scale ecological data such as these are rare, but they are invaluable for testing and improving predictions from vegetation models over

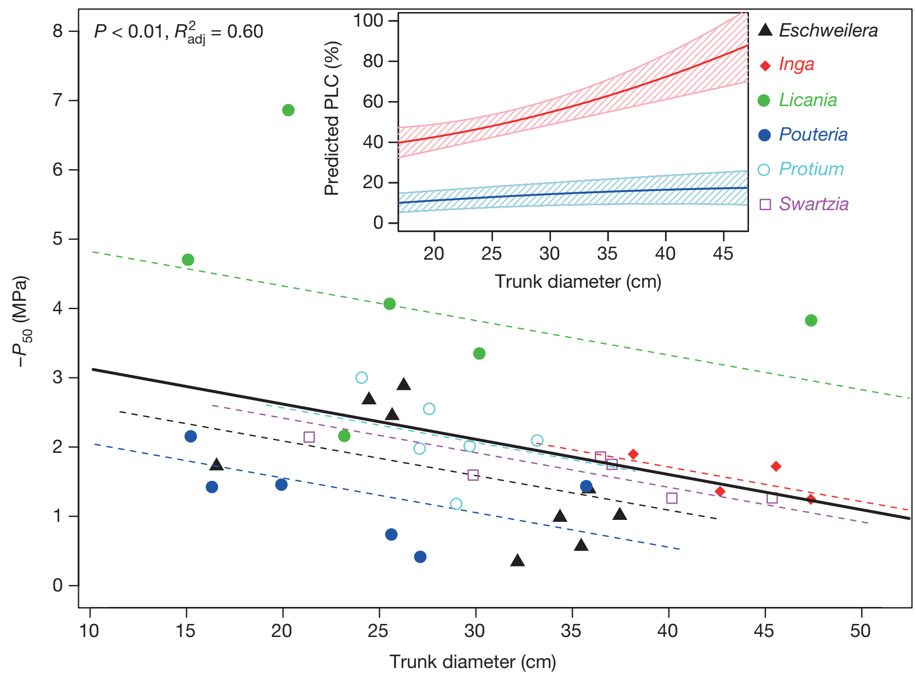


Figure 4 | Xylem vulnerability to embolism and predicted loss of xylem hydraulic conductivity as a function of tree diameter (dbh). A general linear model was employed to test for the effects of genus and tree dbh on the estimates of xylem P_{50} obtained for each species ($n = 37$ trees in total). The dashed lines give the predicted regressions for each of the six genera, whereas the bold black line gives the overall mean regression line across all species (The adjusted R^2 in the figure refers to the overall model, while the P value refers to the significance of tree dbh). The insert shows the predicted losses of hydraulic conductivity (PLC) as a function of tree size and mean midday Ψ_l. In the main graph, each genus is represented with a different symbol and/or colour, as detailed in the legend. In the inset, the bold lines represent the average prediction for the most vulnerable genus (red, *Pouteria*) and for the most resistant one (blue, *Licania*), with the shaded areas (pink, light blue) giving the respective 95% confidence intervals.

timescales that are relevant to climate change[28]. They also underpin the long-term environmental policy needed to manage the natural capital that is embedded in tropical rainforests.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 May; accepted 2 September 2015.
Published online 23 November 2015.

1. Christensen, J. H. *et al.* in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K. *et al.*) (Cambridge Univ. Press, 2013).
2. Mora, C. *et al.* The projected timing of climate departure from recent variability. *Nature* **502,** 183–187 (2013).
3. Reichstein, M. *et al.* Climate extremes and the carbon cycle. *Nature* **500,** 287–295 (2013).
4. Boisier, J. P., Ciais, P., Ducharne, A. & Guimberteau, M. Projected strengthening of Amazonian dry season by constrained climate model simulations. *Nat. Clim. Chang.* **5,** 656–660 (2015).
5. Hartmann, H., Adams, H. D., Anderegg, W. R. L., Jansen, S. & Zeppel, M. J. B. Research frontiers in drought-induced tree mortality: crossing scales and disciplines. *New Phytol.* **205,** 965–969 (2015).
6. da Costa, A. C. L. *et al.* Effect of 7 yr of experimental drought on vegetation dynamics and biomass storage of an eastern Amazonian rainforest. *New Phytol.* **187,** 579–591 (2010).
7. Doughty, C. E. *et al.* Drought impact on forest carbon dynamics and fluxes in Amazonia. *Nature* **519,** 78–82 (2015).
8. O'Brien, M. J., Leuzinger, S., Philipson, C. D., Tay, J. & Hector, A. Drought survival of tropical tree seedlings enhanced by non-structural carbohydrate levels. *Nat. Clim. Chang.* **4,** 710–714 (2014).
9. Nepstad, D. C., Tohver, I. M., Ray, D., Moutinho, P. & Cardinot, G. Mortality of large trees and lianas following experimental drought in an Amazon forest. *Ecology* **88,** 2259–2269 (2007).
10. Phillips, O. L. *et al.* Drought sensitivity of the Amazon rainforest. *Science* **323,** 1344–1347 (2009).
11. Brando, P. M. *et al.* Drought effects on litterfall, wood production and belowground carbon cycling in an Amazon forest: results of a throughfall reduction experiment. *Philos. Trans. R. Soc. B* **363,** 1839–1848 (2008).
12. da Costa, A. C. L. *et al.* Ecosystem respiration and net primary productivity after 8–10 years of experimental through-fall reduction in an eastern Amazon forest. *Plant Ecol. Divers.* **7,** 7–24 (2014).
13. Meir, P. *et al.* Threshold responses to soil moisture deficit by trees and soil in tropical rain forests: insights from field experiments. *Bioscience* **65,** 882–892 (2015).
14. McDowell, N. G. *et al.* The interdependence of mechanisms underlying climate-driven vegetation mortality. *Trends Ecol. Evol.* **26,** 523–532 (2011).
15. Thomas, H. Senescence, ageing and death of the whole plant. *New Phytol.* **197,** 696–711 (2013).
16. Meir, P., Mencuccini, M. & Dewar, R. C. Drought-related tree mortality: addressing the gaps in understanding and prediction. *New Phytol.* **207,** 28–33 (2015).
17. Anderegg, W. R. L. *et al.* Drought's legacy: multiyear hydraulic deterioration underlies widespread aspen forest die-off and portends increased future risk. Glob. *Chang. Biol.* **19,** 1188–1196 (2013).
18. Rowland, L. *et al.* After more than a decade of soil moisture deficit, tropical rainforest trees maintain photosynthetic capacity, despite increased leaf respiration. *Glob. Chang. Biol.* (2015).
19. Anderegg, W. R. *et al.* The roles of hydraulic and carbon stress in a widespread climate-induced forest die-off. *Proc. Natl Acad. Sci. USA* **109,** 233–237 (2012).
20. Choat, B. *et al.* Global convergence in the vulnerability of forests to drought. *Nature* **491,** 752–755 (2012).
21. Sala, A., Woodruff, D. R. & Meinzer, F. C. Carbon dynamics in trees: feast or famine? *Tree Physiol.* **32,** 764–775 (2012).
22. Körner, C. Carbon limitation in trees. *J. Ecol.* **91,** 4–17 (2003).
23. Mencuccini, M. *et al.* Size-mediated ageing reduces vigour in trees. *Ecol. Lett.* **8,** 1183–1190 (2005).
24. Burkhardt, J., Basi, S., Pariyar, S. & Hunsche, M. Stomatal penetration by aqueous solutions – an update involving leaf surface particles. *New Phytol.* **196,** 774–787 (2012).
25. Yates, D. J. & Hutley, L. B. Foliar uptake of water by wet leaves of *Sloanea woollsii*, an australian subtropical rain-forest tree. *Aust. J. Bot.* **43,** 157–167 (1995).
26. Eller, C. B., Lima, A. L. & Oliveira, R. S. Foliar uptake of fog water and transport belowground alleviates drought effects in the cloud forest tree species, *Drimys brasiliensis* (Winteraceae). *New Phytol.* **199,** 151–162 (2013).
27. Fatichi, S., Leuzinger, S. & Körner, C. Moving beyond photosynthesis: from carbon source to sink-driven vegetation modeling. *New Phytol.* **201,** 1086–1095 (2014).
28. Leuzinger, S. *et al.* Do global change experiments overestimate impacts on terrestrial ecosystems? *Trends Ecol. Evol.* **26,** 236–241 (2011).

Acknowledgements This work was supported by UK NERC grant NE/J011002/1 to P.M. and M.M., CNPQ grant 457914/2013-0/MCTI/CNPq/FNDCT/LBA/ESECAFLOR to A.C.L.D., and ARC grant FT110100457 to P.M. It was previously supported by NERC NER/A/S/2002/00487, NERC GR3/11706, EU FP5-Carbonsink and EU FP7-Amazalert to P.M. and J.G., and by grant support to Y.M. from NERC NE/D01025X/1 and the Gordon and Betty Moore Foundation. L.R., M.M. and P.M. would also like to acknowledge support from S. Sitch, Y. Salmon and B. Christoffersen. The authors would also like to thank three anonymous referees for their useful comments.

Author Contributions L.R., P.M., A.C.L.D. and M.M. designed and implemented the research. P.M. conceived and led the experiment and this study. L.R. led recent measurements; all authors contributed to data collection, led by A.C.L.D.; L.R. analysed the data with M.M., P.M., O.J.B. and A.M.P.; L.R. wrote the paper with P.M. and M.M., with contributions from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.R. (lucy.rowland@ed.ac.uk).

METHODS

Site. The through-fall exclusion (TFE) experiment is located in the Caxiuanã National Forest Reserve in the eastern Amazon (1° 43′ S, 51° 27′ W), ~400 km west of the nearest city, Belém, State of Pará, Brazil. The experiment is located in terra firma forest, on yellow oxisol soils which are 75–83% sand, 12–19% clay and 6–10% silt[29]. The site is 15 m above sea level, has a mean annual rainfall between 2,000–2,500 mm and a pronounced dry season between June and November[2].

The experimental site has two 1-ha plots: the TFE, over which plastic panels and gutters have been placed at a height of 1–2 m, and which exclude 50% of the incident rainfall; and a corresponding control plot, $<$50 m from the TFE, on which there has been no manipulation of incident rainfall. The TFE was trenched to between 1–2 m to remove the effect of through-flow of soil water; and to control for any temporary damage to roots from the trenching, the control plot was also trenched to the same depth. The TFE treatment has been installed and running continuously since January 2002 to the present, except for a 1-week period in November 2002 (full removal), a month period during the dry season in November 2014 (sequential removal of all panel) and during 2004 (30% removal).

No statistical methods were used to predetermine sample size, the experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Soil moisture data. In both the control and TFE plots there are soil access pits in which volumetric soil water content sensors (CS616, Campbell Scientific, Logan, USA) located at depths of 0, 0.5, 1, 2.5 and 4 m, monitor soil moisture every hour (see Fisher *et al.*[30] for full methodology). New data presented here are from March 2008–December 2014, averaged into monthly values. Due to equipment failure, some soil moisture data are missing for August and December 2013, for 2008 and 2010 on the control plot, and for November and December 2013 on the TFE plot. Volumetric soil water content was converted in to soil water potential (Ψ_s) using the necessary van Genuchten parameters previously calculated by Fisher *et al.*[31] based on soil hydraulics measurements at this site.

Biomass data. Trees $\geq$10 cm diameter at breast height (dbh) on both the control and the TFE plots were tagged and identified to species level in September 2000. Diameters were measured on these trees at 1.3 m, unless buttress roots were present, in which case the measurement was made above the buttressing. Individuals on the plots were re-censused at varying intervals from January 2001, until November 2014. During each census the trees were also assessed to be either dead or alive. A tree was considered dead if leaflessness was accompanied by a persistent zero or negative stem increment, and/or the tree had snapped or fallen to the ground. Recruitment of new trees into the $\geq$10 cm dbh size class was enumerated in 2005, 2009 and 2014.

Trees in the 10 × 10 m subplots adjacent to the trenches were excluded from our biomass and growth rate analysis to eliminate possible effects of changes in mortality resulting from root damage[2]. Consequently, 369 trees on the control and 358 trees on the TFE, each in 0.64 ha were analysed and the biomass scaled to 1 ha. Following da Costa *et al.*[6] trees were grouped into small (10–20 cm dbh), medium (20–40 cm dbh) and large ($>$40 cm dbh) size classes.

Biomass was calculated using the Chave *et al.*[32] equation which uses diameter, wood density and environmental predictors. A mean and standard deviation of wood density for each tree was calculated from data in the global wood density database[33,34] and from Patiño *et al.*[35]. Multiple estimates of wood density at the species, genus and family level were used to calculate standard deviations on our wood density estimates. Of all the trees on both plots, 68% had values for wood density at species level, 18% at genus level and 3% at family level; 11% of trees were not identified and were given a plot level average wood density with an associated standard deviation. A standard error on our biomass estimations was calculated, which accounted for the error associated with wood density estimation and variations in commonly used allometric equations. Twelve calculations were used to calculate the error on our biomass values; these 12 biomass estimates were from combinations of four biomass equations: Chave *et al.*[32] and Chave *et al.*[36] both with and without height and three wood density estimates: mean wood density and mean wood density ± one standard deviation. The allometric equations selected represent one of the most commonly used biomass equations for Amazonia[36] (with and without height as a predictor variable) and the most recent and most comprehensive biomass equations for our study area[32] (with and without height as a predictor variable). Measurements of height were not available and were calculated from an equation developed specifically for the region of the Amazon in which our plots are located[37]. Mortality rate was calculated according to da Costa *et al.*[6] and separately for trees of 10–20 cm dbh, 20–40 cm dbh and $>$40 cm dbh on the TFE and control plot.

Seasonal growth data. Dendrometers were installed at just above or below the point of dbh measurement on all trees $\geq$10 cm dbh[38]. Circumference measurements from the dendrometer bands were made monthly to tri-monthly from January 2005 to November 2014, with the exception of a six month gap from July 2007 and January 2013. Seasonal shrinkage was calculated by taking the average growth rate of 19 trees on the control and TFE plots which experienced no overall growth, but demonstrated a seasonal pattern of shrinkage and expansion. Tropical trees can experience net diameter shrinkage in the dry season, due to a lack of growth, accompanied by reduced water content of stem tissues, with subsequent swelling when tissues are rehydrated in the wet season[39]. An average pattern of shrinkage and expansion was therefore subtracted from all trees to ensure that dry season growth was not under-estimated and wet season growth over-estimated. Dendrometer increments were filtered to remove the growth spikes from measurement errors following Rowland *et al.*[39]. Subsequent to this procedure, gap-filling was done using linear interpolation[39]; on the control and TFE plot 16% and 17% of the data were gap-filled respectively. Growth rates per day calculated from the dendrometer measurements were averaged into three monthly periods to give continuous tri-monthly growth rates from 2005–2014. When trees in the 10 × 10 m subplots adjacent to the trenches, and trees with poor quality dendrometer measurements throughout the 10 year study period for growth were excluded, 316 and 310 trees on the control and TFE plot remained, respectively. All growth and mortality analyses were done using the R statistical package (version 3.1.2). Tests for significance were performed using Wilcoxon signed-rank test.

Non-structural carbohydrate (NSC) analysis. Using samples cut by a tree climber from fully sunlit branches, three leaves and a branch sample of ~8–10 mm diameter were taken from 20 trees on the control plot and 21 trees on the TFE plot in November 2013, March 2014 and June 2015. The numbers and species of the trees selected for analysis are shown in Extended Data Table 1. The selected trees were all $>$10 cm dbh and represented the most common genera existing on both plots[2]; samples were not taken from the external 10 × 10 m subplots to avoid any impacts of trenching. Samples of tree stem tissue from the same trees were taken with a 5 mm increment borer at breast height once in November 2013; this sampling was not repeated to avoid excessive damage incurred by repeated boring. We followed the enzymatic method proposed by Sevanto *et al.*[40] to analyse the NSC content. Here, NSC is defined as free, low-molecular-weight sugars (glucose, fructose and sucrose) and starch. Immediately after collection, samples were microwaved to stop enzymatic activity. After that, samples were oven-dried at 70 °C for 24–48 h and ground to fine powder. We prepared approximately 12 mg of plant material with 1.6 ml of distilled water for the analysis. We used amyloglucosidase from *Aspergillus niger* (Sigma-Aldrich) to digest total NSC to glucose, and invertase, glucose hexokinase kits (GHK) and phosphorus glucose (Sigma-Aldrich) to quantify the low molecular weight sugars. The concentration of free glucose was determined photometrically in a 96-well microplate spectrophotometer (BioTek, Epoch). NSC values are expressed as per cent of dry matter. For further method details, see Sevanto *et al.*[40].

Xylem vulnerability to cavitation (percentage loss of conductivity; PLC) and leaf water potential measurements. Samples were cut from fully sunlit branches following the same protocol as above for NSC during late dry season (November 2013), and using the same sample trees as used for the NSC analysis (Extended Data Table 1), with the exception of branches of the genus *Manilkara*. To maintain a balance of number of genera and trees sampled, four additional trees were therefore sampled from the genus *Inga* in the TFE and control plots for both sets of analyses. One to three 1.0–1.5 m long branches per tree were cut and left to rehydrate overnight under a black plastic bag in a bucket of water. Maximum vessel length was determined for one branch out of a set of 5–6 branches for each species by injecting low pressure air at the branch base under water and progressively re-cutting the stem until bubbles emerged (maximum conduit length varied between 25 and 50 cm across samples)[41]. Axial slits were made in the branch segments selected for PLC analysis to increase the efficacy of the air injection. These partially debarked segments were mounted on a 4.6 cm long air injection apparatus. Water, filtered to 0.2 µm, flowed gravimetrically through the sample. After 10 min of equilibration at low pressure (10 kPa), the sample was pressurised for 20 min. The pressure was increased in steps of 0.3 to 0.5 MPa, with each step followed by 10 min relaxation and flow measured at a constant background 100 kPa air pressure, until a residual flow lower than 5% of the initial flow was found. Five to ten measurements of water flux were taken at the distal end. The interval for each conductivity measurement ranged from 2 to 10 min depending on stem length, conductivity and pressure head employed (normally 3 kPa). Segment lengths, cross-sectional diameters and leaf areas of the leaves subtended by the measured segment were determined.

We employed a two-parameter Weibull function to model the changes in per cent loss of xylem hydraulic conductivity as a function of xylem pressure[42]. The two parameters represented P_{50} and slope of the conductivity-pressure curve. We estimated P_{50} and slope for all trees using tree as a random factor in a mixed-model analysis (nlme library[43]) in R (Version 3.02, R Core Team). We let P_{50} vary for each tree while keeping slope constant across trees to achieve convergence. We then employed these conditional estimates of P_{50} in a general linear model to test for the effects of plot, genus and dbh (Extended Data Table 2a).

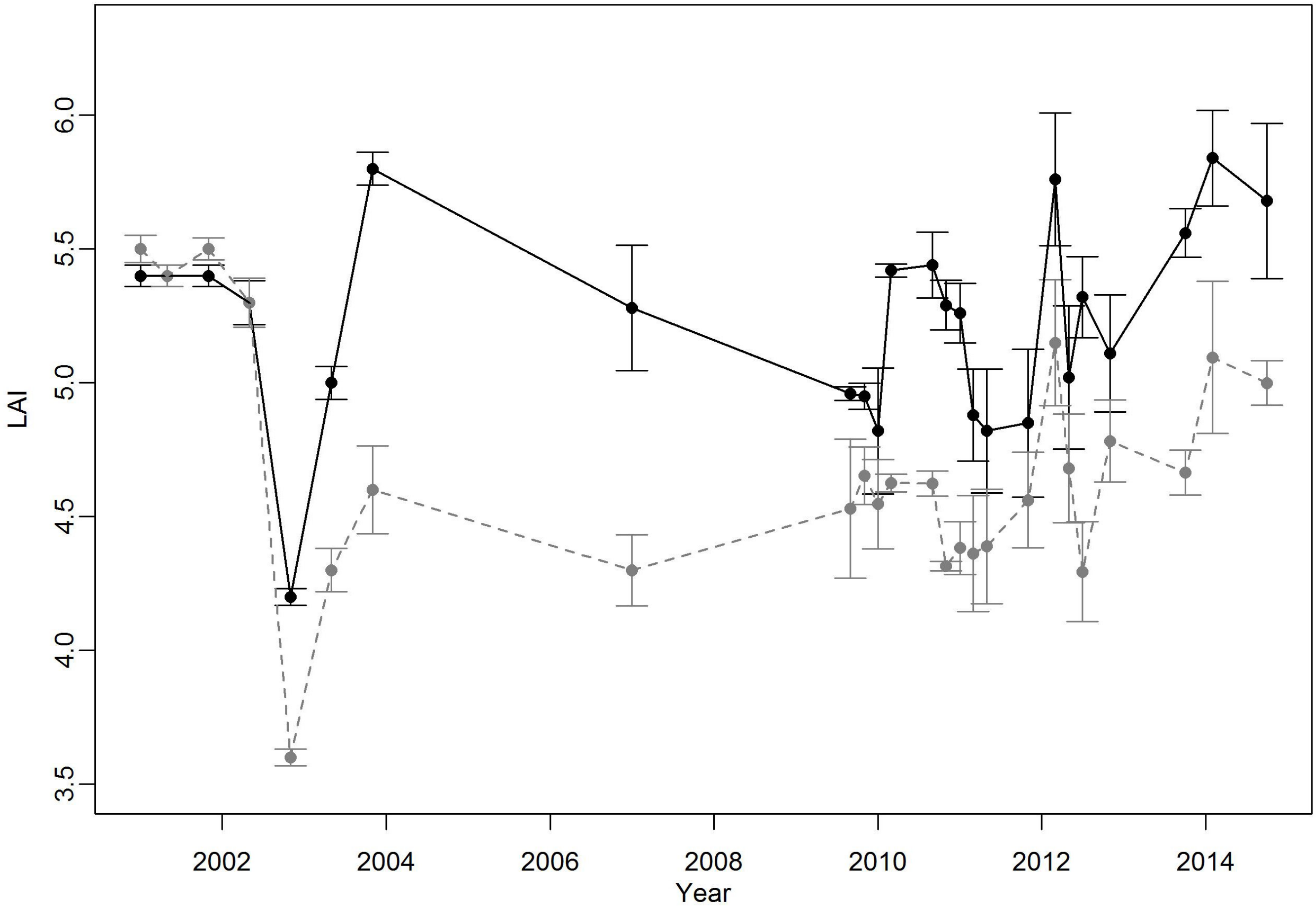
We then confirmed the above results by running a second mixed-effect model, in which we accounted for random tree-by-tree variation and variance driven by phylogeny, by nesting individual trees within genera as the random component of the model and incorporating plot and dbh as fixed effects (Extended Data Table 2b). We tested whether tree diameter affected P_{50} by comparing the performance of the full model with the same model without the effect of dbh on P_{50} using a likelihood ratio test, and by conducting simultaneous hypothesis tests at the 95% significance level (library multcomp in R). For both analyses, distributional assumptions were tested by looking at plots of residuals for fixed and, if the case, random effects.

Leaf water potential Ψ_l was measured during two campaigns (one during the dry and one during the wet season, that is, October 2013 and May 2014, respectively). Two to four leaves were measured for each branch of each tree, following the same sampling protocol above for NSC and hydraulic measurements. During each campaign, Ψ_l measurements were conducted at pre-dawn (before 07:00) and at midday (between 11:30 and 13:00). Midday Ψ_l measurements of both campaigns were used in conjunction with the P_{50} and slope values determined for each genus to estimate percent loss of hydraulic conductivity (PLC), assuming xylem water potential to be equal to measured leaf water potential[20]. We acknowledge that this may overestimate PLC; however in counterpoint, only one seasonal campaign could be conducted and so the minimum of leaf Ψ_l for that year was thus likely to have been underestimated because fuller sampling through the dry season was impractical. In addition, earlier (2003) studies in the same experiment[44] reported minimum values of stem xylem water potentials of around −1.6 MPa, across control and TFE plots, suggesting that xylem values substantially more negative than those assumed here can be experienced.

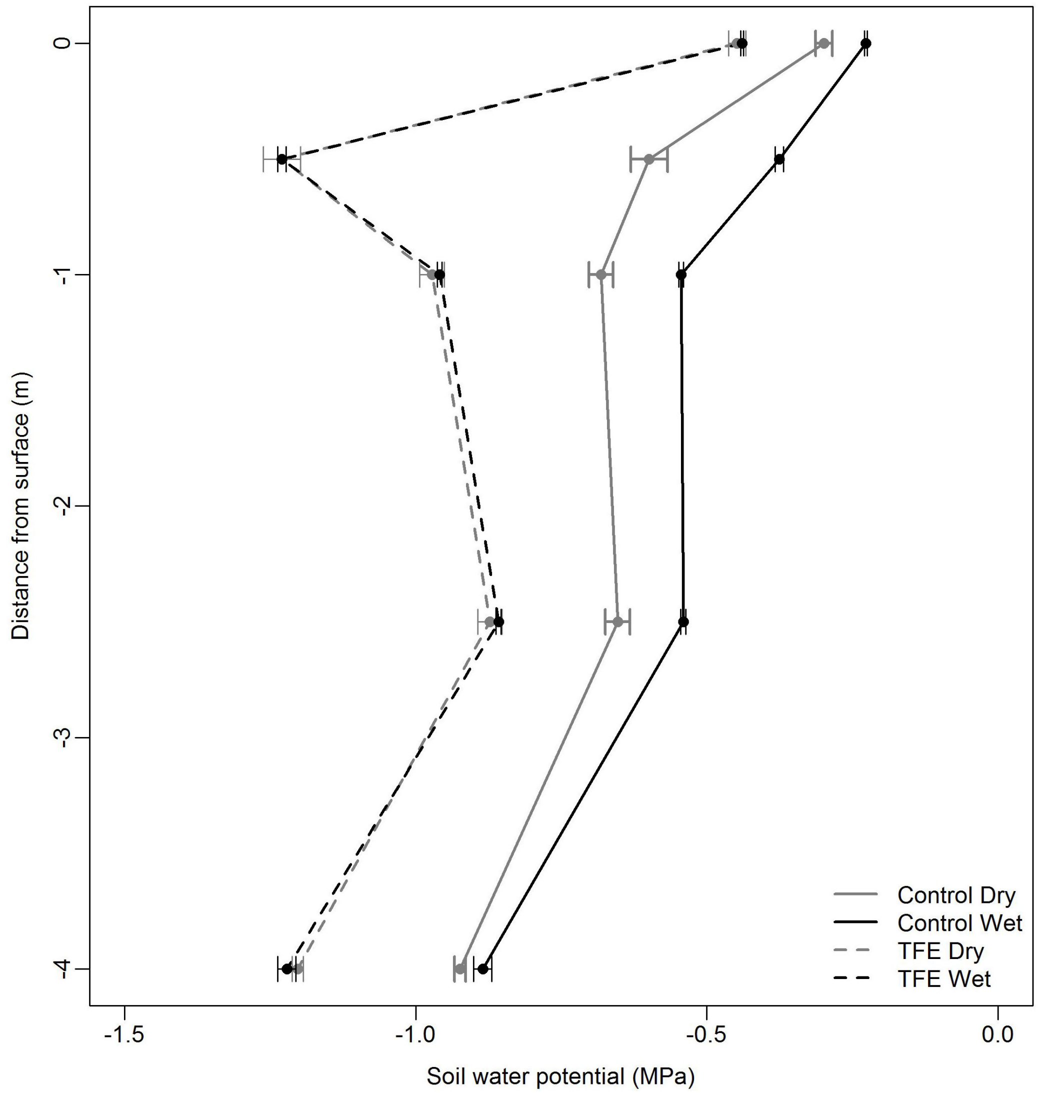
Herbivory. All leaf material from $25 \times 1\,m^2$ litter-traps on the control and TFE was collected 13 times from 2010–2014 at 3–6 month intervals, with one eight month interval in 2011. Each collection of leaf material represented two weeks of litter-fall in the forest. The 7,121 leaves collected from all 25 per plot litter-traps over the study period were scanned and the images were analysed according to Metcalfe *et al.*[45] to calculate the percent leaf area lost to herbivore attack on the control and TFE.

LAI measurements. LAI values from 2001 to 2007 are taken from Fig. 1f in Metcalfe *et al.*[38]. For 2009–2014 LAI was measured at the same 25 permanent points[38] on a grid throughout the control and TFE plot. Measurements were made using hemispherical photos taken every 3–6 months from 2009–2014. Photos were taken before sunrise (~6:00) from a height of 1.5 m on the control plot and 2 m (above the TFE structure) on the TFE plot. The 25 photos per plot were analysed together using the CAN_EYE software (INRA, Avignon, France). Standard errors were calculated using three different estimates of LAI given by the CAN_EYE software.

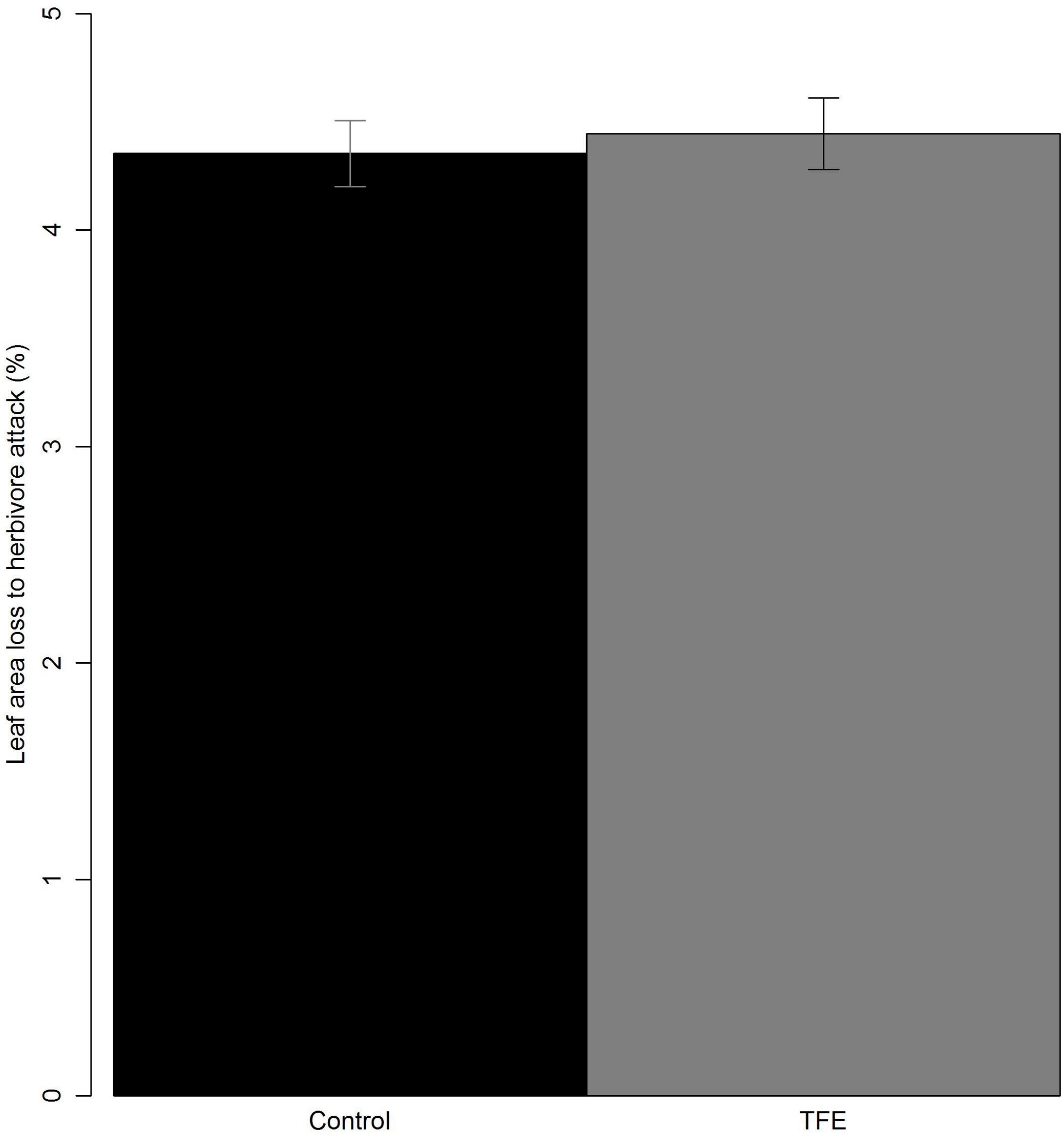
29. Ruivo, M. L. P. & Cunha, E. C. in *Ecosystems and sustainable development IV* (eds Tiezzi, E., Brebbia, C. A. & Uso, J.-L.) Vol. 2, 1113–1121 (WIT Press, 2003).
30. Fisher, R. A. *et al.* The response of an Eastern Amazonian rain forest to drought stress: results and modelling analyses from a throughfall exclusion experiment. *Glob. Chang. Biol.* **13,** 2361–2378 (2007).
31. Fisher, R. A., Williams, M., Ruivo, M. D., de Costa, A. L. & Meira, P. Evaluating climatic and soil water controls on evapotranspiration at two Amazonian rainforest sites. *Agric. For. Meteorol.* **148,** 850–861 (2008).
32. Chave, J. *et al.* Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob. Chang. Biol.* **20,** 3177–3190 (2014).
33. Chave, J. *et al.* Towards a worldwide wood economics spectrum. *Ecol. Lett.* **12,** 351–366 (2009).
34. Zanne, A. E. *et al.* Data from: Towards a worldwide wood economics spectrum. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.234 (2009).
35. Patiño, S. *et al.* Branch xylem density variations across the Amazon Basin. *Biogeosciences* **6,** 545–568 (2009).
36. Chave, J. *et al.* Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* **145,** 87–99 (2005).
37. Feldpausch, T. R. *et al.* Tree height integrated into pantropical forest biomass estimates. *Biogeosciences* **9,** 3381–3403 (2012).
38. Metcalfe, D. B. *et al.* Shifts in plant respiration and carbon use efficiency at a large-scale drought experiment in the eastern Amazon. *New Phytol.* **187,** 608–621 (2010).
39. Rowland, L. *et al.* The sensitivity of wood production to seasonal and interannual variations in climate in a lowland Amazonian rainforest. *Oecologia* **174,** 295–306 (2014).
40. Sevanto, S., McDowell, N. G., Dickman, L. T., Pangle, R. & Pockman, W. T. How do trees die? A test of the hydraulic failure and carbon starvation hypotheses. *Plant Cell Environ.* **37,** 153–161 (2014).
41. Ennajeh, M., Simões, F., Khemira, H. & Cochard, H. How reliable is the double-ended pressure sleeve technique for assessing xylem vulnerability to cavitation in woody angiosperms? *Physiol. Plant.* **142,** 205–210 (2011).
42. Neufeld, H. S. *et al.* Genotypic variability in vulnerability of leaf xylem to cavitation in water-stressed and well-irrigated sugarcane. *Plant Physiol.* **100,** 1020–1028 (1992).
43. Pinheiro, J. C. & Bates, D. M. *Mixed-Effects Models in S and S-PLUS* (Springer, 2000).
44. Fisher, R. A., Williams, M., Do Vale, R. L., Da Costa, A. L. & Meir, P. Evidence from Amazonian forests is consistent with isohydric control of leaf water potential. *Plant Cell Environ.* **29,** 151–165 (2006).
45. Metcalfe, D. B. *et al.* Herbivory makes major contributions to ecosystem carbon and nutrient cycling in tropical forests. *Ecol. Lett.* **17,** 324–332 (2014).



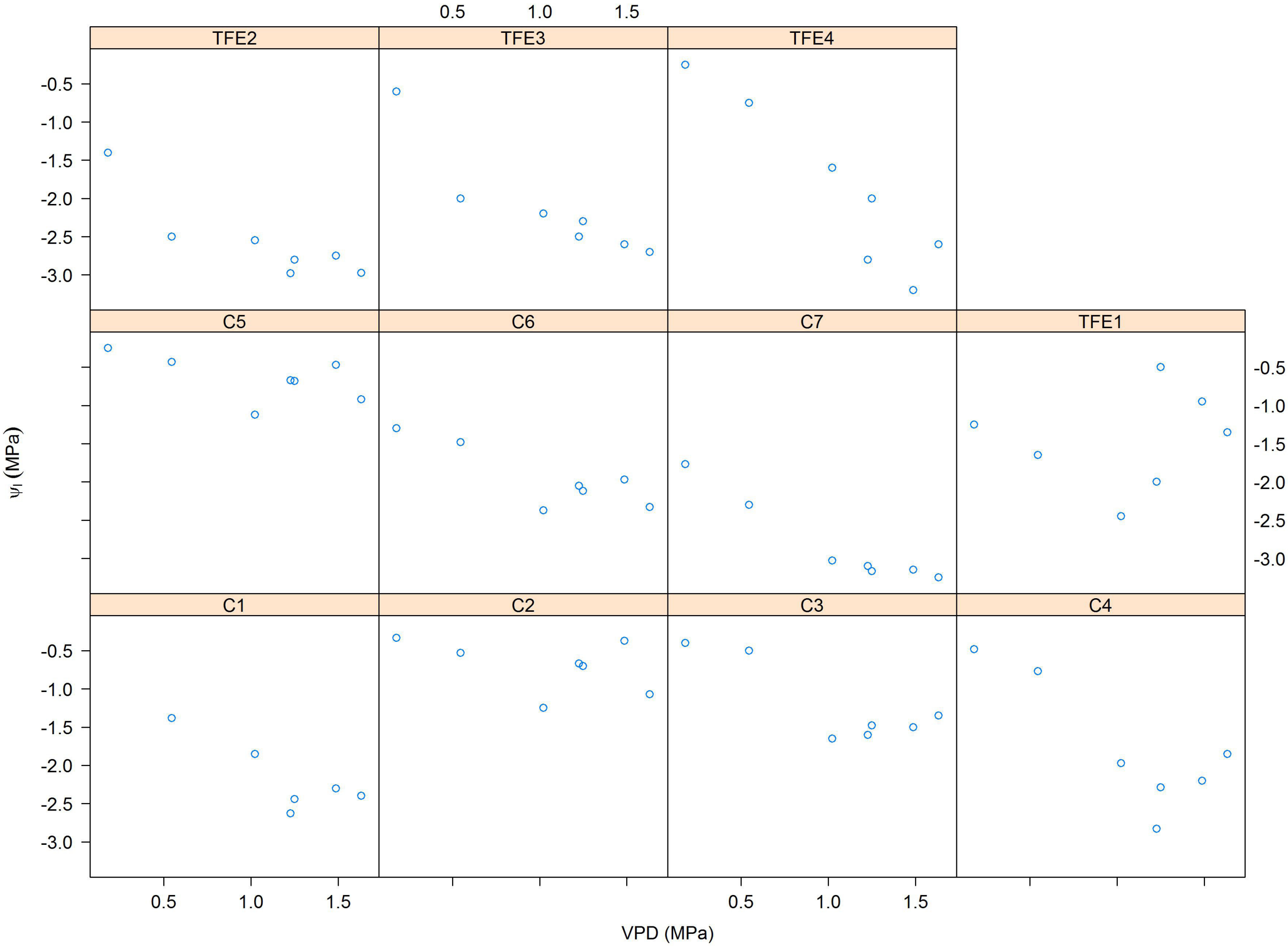
Extended Data Figure 1 | Leaf area index change. Leaf area index (LAI; ratio of leaf area to ground area) for the period of 2001–2014 on the control (black, solid) and TFE (grey, dashed) plots. Error bars show the s.e.m. associated with LAI calculation, which is derived from $n = 25$ photos per control and TFE plot (see Methods).



Extended Data Figure 2 | Seasonal soil water potential. Average soil water potential (−MPa) in the control and TFE during dry season (July–December, control $n = 34$ months, TFE $n = 40$ months) and wet season (January–June control $n = 34$ months, TFE $n = 40$ months), calculated form monthly average volumetric soil moisture content data, collected from 2008–2014, using sensors installed 0, 0.5, 1, 2.5 and 4 m below the surface and the necessary van Genuchten parameters previously calculated from soil hydraulics measurements at this site (see Methods). Error bars show s.e.m.



Extended Data Figure 3 | Leaf herbivory comparison. Average percentage loss of leaf area from herbivore attack calculated from leaves collected in litter-traps on the control ($n = 3{,}297$) and TFE plot ($n = 3{,}824$) from 2010–2014. Error bars show s.e.m and no significant differences were found significant with a $P < 0.05$ using the Wilcoxon test. A separate analysis of herbivore attack on 13,694 top-canopy living leaves from branches of the 41 trees used for the P_{50} analysis support these results, also showing no significant differences in percentage herbivory between the control and the TFE (data not shown).



Extended Data Figure 4 | Diurnal patterns of Ψ_l. Diurnal Ψ_l measured every 2 h from 6:00 until 18:00 in dry season on trees accessible from the walk up tower. Each box shows the diurnal Ψ_l against diurnal air vapour pressure deficit (VPD) from one of seven trees accessible on the control (C), or one of four trees accessible on the TFE. Note that a majority of trees demonstrate an inversely correlated (negative) relationship with VPD. Combined separately for each plot, a significant negative linear relationship is observed between Ψ_l and VPD on the control ($R^2 = 0.18$, $P = 0.002$) and even more strongly on the TFE ($R^2 = 0.33$, $P = 0.001$).

Extended Data Table 1 | NSC and P_{50} sample trees

		Control			TFE		
Genus	**Species**	**Number**	**dbh (cm)**	**P_{50} (MPa)**	**Number**	**dbh (cm)**	**P_{50} (MPa)**
Eschweilera	*grandiflora*	1	36.5	0.6	1	16.8	
		2	34.5	1.0	2	16.7	1.7
					3	24.6	2.7
	coriacea	3	37.6	1.0	4	26.4	2.9
					5	32.3	0.3
					6	36.0	1.4
	pedicellata	4	17.8				
		5	25.8	2.5			
Licania	*membranacea*	6	20.5	6.9	7	15.5	4.7
		7	23.4	2.2	8	25.6	4.1
					9	47.6	3.8
	octandra	8	30.4	3.3			
Manilkara	*bidentata*	9	32.1		10	29.3	
		10	32.5		11	36.7	
		11	48.4		12	56.1	
Pouteria	*anomala*	12	15.4	2.2	13	20.1	1.5
		13	16.4	1.4	14	25.8	0.7
		14	35.9	1.4	15	27.3	0.4
Protium	*tenuifolium*	15	24.3	2.0	16	27.8	2.5
		16	29.9	3.0	17	29.2	1.2
		17	33.4	2.1			
	paniculatum				18	37.3	2.0
Swartzia	*racemosa*	18	30.0	1.6	19	21.5	2.1
		19	37.2	1.3	20	36.2	1.9
		20	45.5	1.7	21	40.3	1.3
Inga	*alba*	21	38.3	1.9	22	45.7	1.7
		22	42.8	1.4	23	47.5	1.2

The genus, species and number of trees > 10 cm dbh sampled for NSC and P_{50} from the central 0.64 ha area of each of the control and TFE plots, and their P_{50} value. The dbh (in cm) of the sample trees is shown in brackets. Where possible, trees from the most common species within the most common genera were sampled; when this was not possible a second species within the same genus was sampled. The genera shown here represent seven of the most abundant genera found across both plots[2]. The samples from the genus *Inga* were only employed for leaf water potential and P_{50} measurements to replace *Manilkara* samples from which P_{50} data were unobtainable.

Extended Data Table 2 | Analysis of the effect of tree dbh on xylem P_{50}

	Degrees of freedom	Sum Squares	Mean Square	F value	Probability	
Tree dbh	1	6.139	6.1390	9.1974	0.0051	**
Plot	1	0.095	0.0950	0.1423	0.7087	ns
Genus	5	32.917	6.5834	9.8632	1.4 e-5	***
Residuals	29	19.357	0.6675			

Model covariates	Degrees of freedom	AIC	Log-likelihood	Likelihood ratio	*P*-value
without tree dbh	6	2793	-1393	28.2	<0.0001
with tree dbh	4	2769	-1378		

Top, analysis of variance table from the general linear model testing for the effects of plot, genus and dbh on xylem P_{50}. Bottom, difference in maximum likelihood for the mixed-model with and without tree size as a predictor of xylem P_{50} (see Methods).
The table shows the main parameters and the significance level of the χ^2 test of the likelihood ratio test. ns, not significant; **, highly significant; ***, very highly significant.

Extended Data Table 3 | Individual tree mortality by genus

control plot			TFE plot		
Genus	**No.**	**Year dead**	**Genus**	**No.**	**Year dead**
Aspidosperma	1	2007	*Abuta*	1	2014
Buchenavia	1	2003	*Bauhinia*	1	2009
Couratari	3	2013,2013,2014	*Brosimum*	1	2005
Duguetia	2	2004,2005	*Chimarrhis*	1	2014
Eschweilera	3	2009, 2013, 2013	*Couepia*	1	2005
Eugenia	1	2001	*Dendrobangia*	2	2005,2005
Franchetella	1	2014	*Derris*	1	2005
Guatteria	2	2005,2005	*Doliocarpus*	1	2009
Helicostylis	1	2007	*Erisma*	1	2006
Inga	2	2009,2013	*Eschweilera*	3	2006,2007,2007
Iryanthera	1	2013	*Forsteronia*	1	2009
Licania	4	2002,2008,2013,2013	*Goupia*	4	2002,2004,2013,2014
Micropholis	1	2002	*Guatteria*	2	2008,2009
Minquartia	2	2003,2004	*Hirtella*	1	2009
NI	4	2001,2002,2013,2013	*Inga*	4	2005,2008,2008,2014
Parkia	1	2014	*Iryanthera*	1	2014
Pouteria	4	2003,2007,2009,2010	*Lecythis*	5	2004,2009,2009,2013,2013
Protium	4	2005,2009,2014,2014	*Licania*	2	2008, 2009
Rinoria	4	2004,2009,2013,2014	*Machaerium*	1	2003
Sclerolobium	1	2005	*Manilkara*	4	2004,2004,2008,2014
Stachyarrhena	1	2014	*Marmaroxylon*	1	2001
Stryphnodendron	1	2006	*Mezilaurus*	1	2008
Swartzia	4	2002,2003,2007,2014	*Micropholis*	4	2003,2005,2007,2009
Tapura	1	2014	*Minquartia*	1	2009
Tetragastris	2	2006,2013	*Naucleopsis*	1	2002,2007,2014
Vantanea	1	2003	*Newtonia*	1	2009
Virola	1	2006	*NI*	2	2002,2013
Vouacapoua	2	2013,2014	*Ocotea*	4	2002,2008,2009,2013
			Oenocarpus	1	2008
			Ormosia	2	2001, 2009
			Ouratea	4	2006,2007,2009,2014
			Pouteria	6	2005,2005,2007, 2009,2013,2014
			Pradosia	1	2005
			Protium	5	2005,2007,2007,2009,2009
			Pseudolmedia	2	2002, 2006
			Quararibea	1	2009
			Sclerolobium	1	2014
			Stachyarrhena	3	2007,2013,2014
			Swartzia	1	2004
			Symphonia	1	2014
			Tetragastris	1	2006, 2013, 2014
			Xylopia	5	2002,2004,2005,2014,2014

The number (No.) of dead trees per genus and the year of death (Year dead) for trees on the control and the TFE, excluding the outer subplots (see Methods). Trees not identified to genus are shown as *NI*.

LETTER

doi:10.1038/nature16159

Reversal of phenotypes in *MECP2* duplication mice using genetic rescue or antisense oligonucleotides

Yehezkel Sztainberg[1,2], Hong-mei Chen[3,4,5], John W. Swann[3,4,5], Shuang Hao[2,5], Bin Tang[2,5], Zhenyu Wu[2,5], Jianrong Tang[2,5], Ying-Wooi Wan[2,6], Zhandong Liu[2,5], Frank Rigo[7] & Huda Y. Zoghbi[1,2,5,8]

Copy number variations have been frequently associated with developmental delay, intellectual disability and autism spectrum disorders[1]. *MECP2* duplication syndrome is one of the most common genomic rearrangements in males[2] and is characterized by autism, intellectual disability, motor dysfunction, anxiety, epilepsy, recurrent respiratory tract infections and early death[3–5]. The broad range of deficits caused by methyl-CpG-binding protein 2 (MeCP2) overexpression poses a daunting challenge to traditional biochemical-pathway-based therapeutic approaches. Accordingly, we sought strategies that directly target MeCP2 and are amenable to translation into clinical therapy. The first question that we addressed was whether the neurological dysfunction is reversible after symptoms set in. Reversal of phenotypes in adult symptomatic mice has been demonstrated in some models of monogenic loss-of-function neurological disorders[6–8], including loss of MeCP2 in Rett syndrome[9], indicating that, at least in some cases, the neuroanatomy may remain sufficiently intact so that correction of the molecular dysfunction underlying these disorders can restore healthy physiology. Given the absence of neurodegeneration in *MECP2* duplication syndrome, we propose that restoration of normal MeCP2 levels in *MECP2* duplication adult mice would rescue their phenotype. By generating and characterizing a conditional *Mecp2*-overexpressing mouse model, here we show that correction of MeCP2 levels largely reverses the behavioural, molecular and electrophysiological deficits. We also reduced MeCP2 using an antisense oligonucleotide strategy, which has greater translational potential. Antisense oligonucleotides are small, modified nucleic acids that can selectively hybridize with messenger RNA transcribed from a target gene and silence it[10,11], and have been successfully used to correct deficits in different mouse models[12–18]. We find that antisense oligonucleotide treatment induces a broad phenotypic rescue in adult symptomatic transgenic *MECP2* duplication mice (*MECP2*-TG)[19,20], and corrected *MECP2* levels in lymphoblastoid cells from *MECP2* duplication patients in a dose-dependent manner.

To determine whether *MECP2* duplication syndrome is reversible, we generated a conditional *MECP2* overexpression mouse model that carries two functional alleles with species-matched endogenous control elements: a human wild-type *MECP2* allele, and a conditional mouse *Mecp2* allele (*Mecp2^{lox}*) that can be deleted using tamoxifen-inducible Cre recombination (Fig. 1a). Our breeding strategy resulted in FVB/N × C57Bl/6 F1 hybrid mice belonging to the following three genotypes: Flox, Flox;TG and Flox;TG;Cre. The *loxP* sequences did not alter MeCP2 expression or phenotype as Flox and Flox;TG mice were indistinguishable from wild-type and transgenic (TG) mice, respectively, in both molecular and behavioural assays (Extended Data Figs 1 and 2). To ascertain the efficiency of Cre-mediated recombination, we injected Flox;TG;Cre mice intraperitoneally with either tamoxifen (TMX) or vehicle over the course of 4 weeks (Fig. 1b), and euthanized four cohorts of mice at different time points after initiation of treatment. MeCP2 protein levels were significantly downregulated at 2.5 weeks, and the levels of MeCP2 remained low thereafter (Fig. 1c, d).

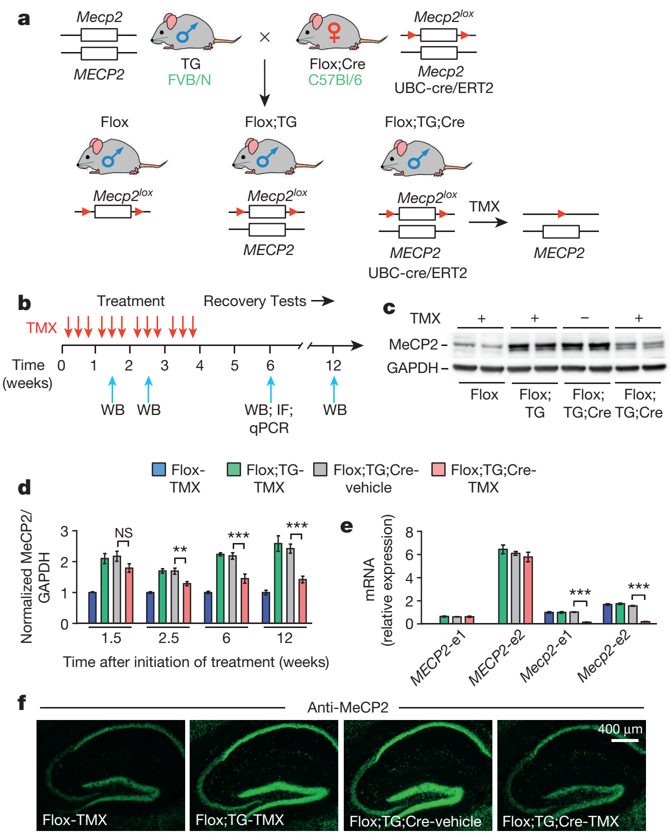


Figure 1 | Inducible Cre–*lox* recombination normalizes MeCP2 levels in adult *MECP2* duplication mice. a, Breeding strategy to generate conditional *MECP2* overexpression mice. To mediate recombination, we used a Cre recombinase driven by a ubiquitin C promoter and fused to a modified human oestrogen receptor (UBC-cre/ERT2). **b**, Tamoxifen (TMX) treatment protocol and the time points for western blot (WB), immunofluorescence (IF) and RT–qPCR. **c**, Western blot from cortical samples at 6 weeks (for gel source data, see Supplementary Fig. 1). **d**, Kinetics of MeCP2 levels ($n = 6$; for gel source data, see Supplementary Fig. 1). **e**, RT–qPCR from cortical samples with specific primers for human or mouse *Mecp2*, and for each of the two alternatively spliced isoforms ($n = 6$). **f**, Immunostaining for MeCP2 in hippocampal slices. NS, not significant. Data are mean ± s.e.m. $^{**}P < 0.01$; $^{***}P < 0.001$ (two-tailed *t*-test).

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA. [2]Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, Texas 77030, USA. [3]The Cain Foundation Laboratories, Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, Texas 77030, USA. [4]Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, USA. [5]Department of Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA. [6]Department of Obstetrics and Gynecology, Baylor College of Medicine, Houston, Texas 77030, USA. [7]Isis Pharmaceuticals, 2855 Gazelle Court, Carlsbad, California 92010, USA. [8]Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA.

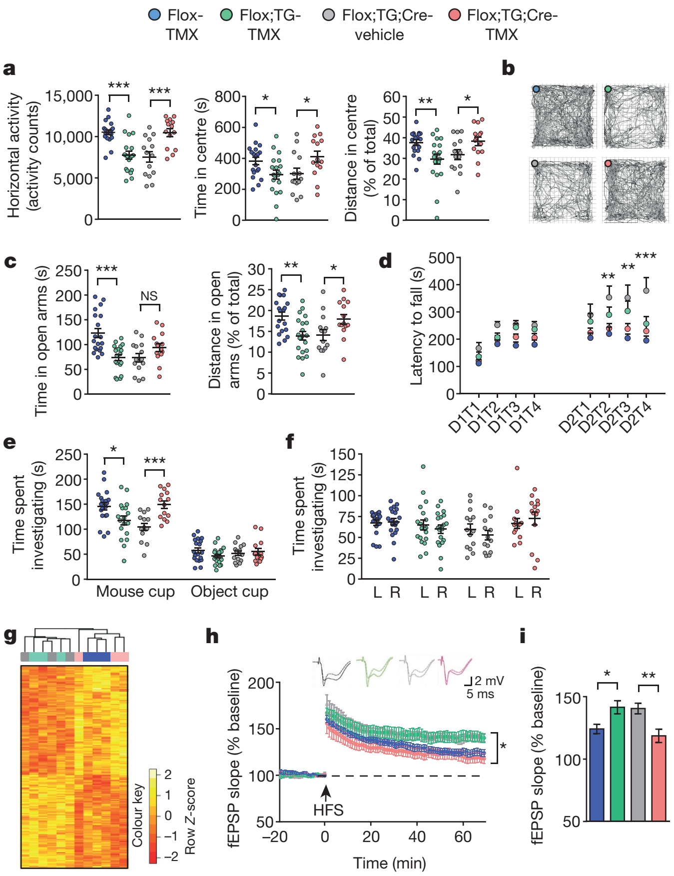


Figure 2 | Genetic normalization of MeCP2 levels reverses deficits in adult *MECP2* duplication mice. a, Reversal of hypoactivity and anxiety-like behaviours in the open field. **b**, Representative tracking plots of the open field test. **c**, Reversal of anxiety-like behaviour in the elevated plus maze test. **d**, Reversal of abnormal motor behaviour on the rotarod test (asterisks indicate significant difference between Flox;TG;Cre-TMX and Flox;TG;Cre-vehicle groups). D1T1, day 1, trial 1. **e**, Reversal of impaired social behaviour in the 3-chamber test. **f**, No preference for the left (L) or right (R) chambers in the habituation phase of the test. $n = 19$ for Flox-TMX and Flox;TG-TMX groups; $n = 14$ for Flox;TG;Cre-vehicle and Flox;TG;Cre-TMX groups. **g**, Transcriptional heat map for the hippocampus ($n = 3$). **h**, TMX treatment normalized LTP in Flox;TG;Cre mice ($n = 6$–7 mice, 15–20 slices). Top, representative electrophysiological traces at baseline and after high-frequency stimulation (HFS). **i**, Quantification of the last 10 min of the LTP recording ($n = 6$–7 mice, 15–20 slices). Data are mean ± s.e.m. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$ (two-tailed *t*-test (**a**, **c**, **e**, **f**, **i**), and repeated-measures two-way analysis of variance (ANOVA) followed by Tukey honest significant difference (HSD) post hoc correction for multiple comparisons (**d**, **h**)).

Moreover, quantitative reverse transcription PCR (RT–qPCR) showed that Cre-mediated recombination efficiently downregulated mRNA levels of both alternatively spliced isoforms (*Mecp2*-e1 and *Mecp2*-e2) of the floxed mouse *Mecp2*, but not the human transgenic *MECP2* allele (Fig. 1e). Finally, we confirmed the normalization of MeCP2 levels by immunofluorescence staining of hippocampal slices (Fig. 1f).

Next, we injected a new cohort of 8–9-week-old mice with TMX or vehicle for behavioural characterization. Flox;TG;Cre mice injected with TMX (Flox;TG;Cre-TMX) were indistinguishable from Flox control mice in the different assays, showing a resolution of the phenotypes that resemble *MECP2* duplication syndrome, such as hypoactivity, anxiety-like behaviour, motor abnormalities and social behaviour deficits (Fig. 2a–f).

Changes in MeCP2 abundance affect the mRNA levels of thousands of genes in the brain[20–22]. Therefore, we proposed that normalizing MeCP2 levels would also normalize gene expression patterns. First, we analysed the expression of selected MeCP2-sensitive genes in the hypothalamus[21] and the cerebellum[22] by RT–qPCR, in adult mice. The mRNA levels of these genes in the Flox;TG;Cre-TMX group were indistinguishable from the Flox control group (Extended Data Fig. 3a, b). We then performed whole transcriptome sequencing (RNA-seq) analysis to evaluate expression patterns in the hippocampus. The analysis showed that the mRNA expression profile of the Flox;TG;Cre-TMX group clustered together with the Flox control group (Fig. 2g and Supplementary Table 1). Reducing MeCP2 to normal levels in symptomatic mice thus seems to rescue the behavioural phenotype by reversing pathogenic molecular changes in the brain.

MeCP2 levels also influence synaptic plasticity, as indicated by abnormalities in hippocampal long-term potentiation (LTP) in MeCP2-null[23] and *MECP2*-TG mice[19]. We therefore assessed LTP at the Schaffer collateral synapses of the CA1 area in the hippocampus, and found no difference in the input–output relationship or paired-pulse facilitation between the groups, indicating that MeCP2 overexpression does not affect basal synaptic transmission or short-term plasticity (Extended Data Fig. 4a–d). However, normalization of MeCP2 levels in Flox;TG;Cre-TMX mice completely rescued the abnormal, enhanced LTP (Fig. 2h, i). Defects in long-term hippocampal synaptic plasticity induced by MeCP2 overexpression are thus reversible in adult mice.

To determine whether we could normalize MeCP2 levels using a strategy that is more readily translatable into a medical therapy, we took advantage of our *MECP2*-TG mice containing one copy of human *MECP2* (in addition to the endogenous mouse gene) to screen for a treatment using human-specific antisense oligonucleotides (ASOs). We tested ASOs designed to bind several regions of the human *MECP2* precursor mRNA (pre-mRNA) so as to reduce the levels of both alternatively spliced *MECP2* isoforms, *MECP2*-e1 and *MECP2*-e2 (Extended Data Fig. 5a). After screening *MECP2* ASOs for their ability to reduce *MECP2* levels in cultured human cells, and for toxicity in wild-type mice (data not shown), we screened five selected *MECP2* ASOs by injecting them stereotaxically into the brain of *MECP2*-TG mice (Extended Data Fig. 5 and Extended Data Table 1), and used the most effective one, ASO-5, for further studies.

To determine the duration of treatment efficacy, we gradually infused ASO into the brains of 7–8-week-old mice using micro-osmotic pumps designed to deliver the molecule at a constant rate over a 4-week period (Fig. 3a and Extended Data Fig. 6). At the end of treatment, immunofluorescence staining showed that the ASO was widely distributed throughout the brain (Fig. 3b) and that it effectively knocked down MeCP2 to close to wild-type levels (Fig. 3c). We next analysed MeCP2 expression in the cortex at different time points after initiation of treatment by western blot (as described in Fig. 3a). MeCP2 was significantly downregulated 4 weeks after the initiation of the treatment (Fig. 3d, e), and remained so for an additional 4 weeks after stopping the infusion (Fig. 3e). We further confirmed the specificity of the ASO for human *MECP2* by RT–qPCR (Fig. 3f).

We then treated a new cohort of animals for behavioural characterization. At 6–7 weeks after the initiation of the treatment, rescue was evident only in the rotarod test (Fig. 3i and Extended Data Fig. 7), but by 10–11 weeks, the hypoactivity, anxiety-like behaviour and social behaviour of *MECP2*-TG mice were also reversed (Fig. 3g, h, j). Ten weeks after treatment cessation, when MeCP2 levels had increased to pre-treatment levels, the symptoms reappeared (data not shown). To gain insight into the basis of the delayed behavioural rescue, we performed RNA-seq analysis in the hippocampus at two different time points after treatment initiation (Fig. 4a and Supplementary Table 2). We found that 4 weeks after initiation of treatment, there was a trend towards normalization of the expression of some mRNAs, but the transgenic-ASO group did not cluster together with the wild-type group. By 8 weeks, however, the transgenic-ASO group clustered together with the wild-type group, suggesting that reversal of pathogenic molecular changes in the brain as a consequence

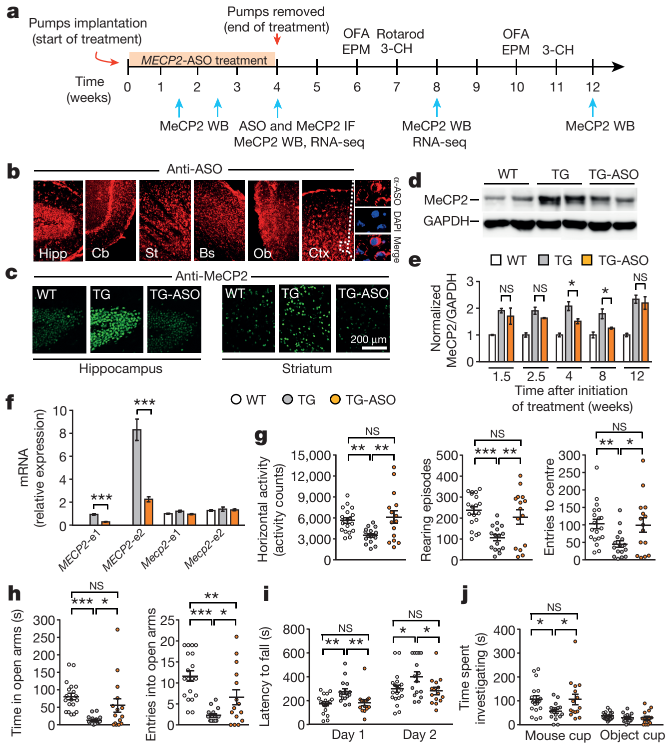


Figure 3 | Gradual infusion of ASO normalizes MeCP2 levels and reverses abnormal behaviour. a, Timeline of gradual ASO treatment and western blot, immunofluorescence, RNA-seq and behavioural tests. 3-CH, 3 chamber; EPM, elevated plus maze; OFA, open field activity. **b**, Immunofluorescence staining against ASOs. Bs, brainstem; cb, cerebellum; ctx, cortex; hipp, hippocampus; ob, olfactory bulb; st, striatum. **c**, Immunofluorescence staining for MeCP2. **d**, Representative western blot of MeCP2 in the cortex 4 weeks after initiation of treatment (for gel source data, see Supplementary Fig. 2). **e**, Kinetics of MeCP2 levels ($n = 4$–5; for gel source data, see Supplementary Fig. 2). **f**, RT–qPCR from cortical samples with specific primers for mouse or human *MECP2* and for each of the two alternatively spliced *MECP2* isoforms ($n = 3$). **g**, Reversal of hypoactivity and anxiety-like behaviour in the open field at week 10. **h**, Reversal of anxiety-like behaviour in the elevated plus maze test at week 10. **i**, Reversal of abnormal motor behaviour on the rotarod test at week 7. Data were averaged per day over four trials. **j**, Reversal of abnormal social behaviour in the 3-chamber test at week 11. $n = 19$ for wild-type group; $n = 16$ for transgenic (TG) group; $n = 15$ for TG-ASO group. Data are mean ± s.e.m. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$ (two-tailed t-test (**e**, **f**), and one-way ANOVA followed by Fisher's least significant difference (LSD) post hoc test (**g**–**j**)).

of MeCP2 normalization correlates strongly with resolution of the behavioural phenotype. To test for possible off-target effects, we compared expression profiles at both time points to detect genes whose expression was significantly affected by the ASO treatment (transgenic versus transgenic-ASO), but was not different between wild-type and transgenic mice. We found only 10 overlapping genes that meet this criterion (see Supplementary Table 2), suggesting minimal off-target effects.

Seizures occur in *MECP2* duplication syndrome mice as they age[19], so we tested the ability of ASO treatment to reverse abnormal electrographic discharges in 25–35-week-old *MECP2*-TG1 mice (pure FVB/N background). Vehicle-treated *MECP2*-TG1 mice manifested electrographic seizure spikes in electroencephalography (EEG) recordings from the cortex (Fig. 4b and Extended Data Fig. 8), and strong electrographic seizure events were typically accompanied by behavioural seizures (Supplementary Video 1). ASO treatment abolished these abnormal EEG discharges and eliminated behavioural seizures and electrographic seizure spikes in this group (Fig. 4b and Extended Data Fig. 8).

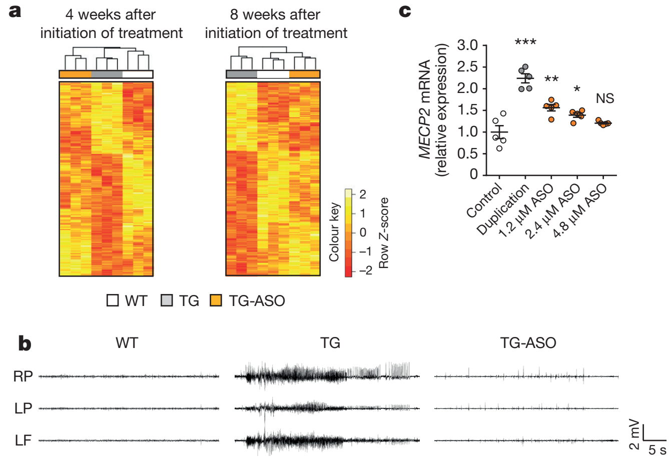


Figure 4 | ASO treatment corrects abnormal gene expression and EEG. a, Transcriptional heat maps of the hippocampus ($n = 3$ mice per group). **b**, Representative EEG traces ($n = 3$ for wild-type group; $n = 4$ for transgenic group, $n = 6$ for transgenic-ASO group). LF, left frontal cortex; LP, left parietal cortex; RP, right parietal cortex. **c**, ASO-corrected *MECP2* mRNA levels in lymphoblastoid cells from *MECP2* duplication patients ($n = 5$). Asterisks denote statistical difference to the control group. Data are mean ± s.e.m. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$ (one-way ANOVA followed by Tukey HSD post hoc correction for multiple comparisons).

Lastly, we determined that *MECP2*-ASO treatment corrected *MECP2* mRNA levels in a dose-dependent manner in lymphoblastoid cells from *MECP2* duplication patients (Fig. 4c and Extended Data Table 2). The amount of ASO can thus be titrated and optimized so as to target only some *MECP2* RNA and allow translation of physiological levels of the protein.

As the main 'reader' of methylated cytosines[24], MeCP2 has a fundamental role in epigenetics, controlling chromatin states and the expression of thousands of genes[21,25,26]. Accordingly, MeCP2 expression must be maintained within a fairly narrow range to assure proper gene expression and neuronal function[27]. Here we demonstrated that restoration of MeCP2 to its normal level can largely reverse the phenotype of adult symptomatic *MECP2* duplication mice.

It is worth noting that reversal of *MECP2* duplication-like features was evident 6–7 weeks after genetic rescue, but took 10–11 weeks after initiation of ASO treatment. This 4-week difference is probably due to the more gradual reduction in MeCP2 levels brought about by ASO treatment. The RNA-seq results in the ASO experiment correlated well with the resolution of the disease phenotype: abnormal gene expression was still prominent 4 weeks after the initiation of treatment (Fig. 4a), whereas robust correction of gene expression 8 weeks after treatment initiation (and 4 weeks after MeCP2 levels normalized (Fig. 3e)) was accompanied by full behavioural rescue. In mice, therefore, MeCP2 levels must be normal for 1 month before there is resolution of the duplication phenotypes.

The finding that ASO treatment rescues the *MECP2* duplication-like phenotypes to a similar extent as the genetic rescue provides a proof-of-concept about the value of this approach. To move this closer to translation, further studies will have to test different ASO dosages and establish the safety margin of MeCP2 levels, using a mouse model that exclusively expresses two human *MECP2* alleles. Additionally, we will screen thousands of *MECP2* ASOs for off-target effects.

Overall, our results show that delivering ASOs to the central nervous system is a promising therapeutic approach for treating *MECP2* duplication syndrome, and has potential for other disorders caused by duplication of genetic material by targeting genes in the respective critical regions, such as peripheral myelin protein 22 (*PMP22*) in Charcot–Marie–Tooth disease[28], retinoic acid induced receptor 1 (*RAI1*) in Potocki–Lupski syndrome[29], and dual specificity tyrosine-phosphorylation-regulated kinase 1A (*DYRK1A*) in Down syndrome[30].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 December 2014; accepted 19 October 2015.

Published online 25 November 2015.

1. Malhotra, D. & Sebat, J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* **148,** 1223–1241 (2012).
2. Lugtenberg, D. *et al.* Structural variation in Xq28: MECP2 duplications in 1% of patients with unexplained XLMR and in 2% of male patients with severe encephalopathy. *Eur. J. Hum. Genet.* **17,** 444–453 (2009).
3. Van Esch, H. *MECP2* duplication syndrome. *Mol. Syndromol*. **2,** 128–136 (2011).
4. Ramocki, M. B., Tavyev, Y. J. & Peters, S. U. The MECP2 duplication syndrome. *Am. J. Med. Genet. A.* **152A,** 1079–1088 (2010).
5. Lubs, H. *et al.* XLMR syndrome characterized by multiple respiratory infections, hypertelorism, severe CNS deterioration and early death localizes to distal Xq28. *Am. J. Med. Genet.* **85,** 243–248 (1999).
6. Costa, R. M. *et al.* Mechanism for the learning deficits in a mouse model of neurofibromatosis type 1. *Nature* **415,** 526–530 (2002).
7. Ehninger, D. *et al.* Reversal of learning deficits in a $Tsc2^{+/-}$ mouse model of tuberous sclerosis. *Nature Med.* **14,** 843–848 (2008).
8. Dölen, G. *et al.* Correction of fragile X syndrome in mice. *Neuron* **56,** 955–962 (2007).
9. Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. Reversal of neurological defects in a mouse model of Rett syndrome. *Science* **315,** 1143–1147 (2007).
10. Southwell, A. L., Skotte, N. H., Bennett, C. F. & Hayden, M. R. Antisense oligonucleotide therapeutics for inherited neurodegenerative diseases. *Trends Mol. Med.* **18,** 634–643 (2012).
11. Bennett, C. F. & Swayze, E. E. RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu. Rev. Pharmacol. Toxicol.* **50,** 259–293 (2010).
12. Lentz, J. J. *et al.* Rescue of hearing and vestibular function by antisense oligonucleotides in a mouse model of human deafness. *Nature Med.* **19,** 345–350 (2013).
13. Wheeler, T. M. *et al.* Targeting nuclear RNA for *in vivo* correction of myotonic dystrophy. *Nature* **488,** 111–115 (2012).
14. Hua, Y. *et al.* Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model. *Nature* **478,** 123–126 (2011).
15. Meng, L. *et al.* Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* **518,** 409–412 (2015).
16. Kordasiewicz, H. B. *et al.* Sustained therapeutic reversal of Huntington's disease by transient repression of huntingtin synthesis. *Neuron* **74,** 1031–1044 (2012).
17. Passini, M. A. *et al.* Antisense oligonucleotides delivered to the mouse CNS ameliorate symptoms of severe spinal muscular atrophy. *Sci. Transl. Med.* **3,** 72ra18 (2011).
18. Smith, R. A. *et al.* Antisense oligonucleotide therapy for neurodegenerative disease. *J. Clin. Invest.* **116,** 2290–2296 (2006).
19. Collins, A. L. *et al.* Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Hum. Mol. Genet.* **13,** 2679–2689 (2004).
20. Samaco, R. C. *et al. Crh* and *Oprm1* mediate anxiety-related behavior and social approach in a mouse model of MECP2 duplication syndrome. *Nature Genet.* **44,** 206–211 (2012).
21. Chahrour, M. *et al.* MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320,** 1224–1229 (2008).
22. Ben-Shachar, S., Chahrour, M., Thaller, C., Shaw, C. A. & Zoghbi, H. Y. Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. *Hum. Mol. Genet.* **18,** 2431–2442 (2009).
23. Chao, H. T. *et al.* Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* **468,** 263–269 (2010).
24. Lewis, J. D. *et al.* Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69,** 905–914 (1992).
25. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393,** 386–389 (1998).
26. Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genet.* **19,** 187–191 (1998).
27. Chao, H. T. & Zoghbi, H. Y. MeCP2: only 100% will do. *Nature Neurosci.* **15,** 176–177 (2012).
28. Patel, P. I. *et al.* The gene for the peripheral myelin protein PMP-22 is a candidate for Charcot-Marie-Tooth disease type 1A. *Nature Genet.* **1,** 159–165 (1992).
29. Potocki, L. *et al.* Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. *Am. J. Hum. Genet.* **80,** 633–649 (2007).
30. Arron, J. R. *et al.* NFAT dysregulation by increased dosage of DSCR1 and DYRK1A on chromosome 21. *Nature* **441,** 595–600 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Jaenisch for the *Mecp2*^{tm1Jae} mice, S. Chun for ASOs tolerability studies in wild-type mice, L. Lombardi, M. Rousseaux, C. Alcott and V. Brandt for critical input, and C. Spencer for behavioural assays training. We are indebted to the patients and families who participated in this study. This project was funded by the National Institutes of Health (5R01NS057819 and 5P30HD024064 to H.Y.Z.), the Rett Syndrome Research Trust (401 Project), the Carl. C. Anderson, Sr and Marie Jo Anderson Charitable Foundation, the Howard Hughes Medical Institute (H.Y.Z.), NSF DMS-1263932 (Z.L.), and the Baylor Intellectual Disabilities Research Center (1U54HD083092) neurovisualization, neuroconnectivity and neurobehavioral cores.

Author Contributions Y.S. performed the behavioural and molecular experiments. Y.S., H.C., J.W.S., J.T., F.R. and H.Y.Z. contributed to the concept and design of the experiments. H.C. performed the electrophysiological (LTP) experiments. S.H., Z.W. and B.T. performed the EEG experiments. Y.S., H.C., J.W.S., F.R., Y.-W.W., Z.L., S.H., J.T. and H.Y.Z. collected, analysed and interpreted the data. Y.S. and H.Y.Z. wrote and edited the paper.

Author Information RNA-seq data have been deposited in the Gene Expression Omnibus under accession number GSE71235. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.Y.Z. (hzoghbi@bcm.edu).

METHODS

ASO synthesis. Isis Pharmaceuticals synthesized all ASOs as previously described[15]. All ASOs consist of 20 chemically modified nucleotides (MOE gapmer). The central gap of 10 deoxynucleotides is flanked on its 5′ and 3′ sides by five 2′-*O*-(2-methoxyethyl) (MOE)-modified nucleotides. The backbone modifications from 5′ to 3′ are: 1-PS, 4-PO, 10-PS, 2-PO and 2-PS. Phosphorothioate (PS) modifications were replaced with native phosphodiester (PO) in the MOE wings to reduce the overall PS content of the ASO, since a fully modified PS ASO is not necessary for robust CNS activity[15]. The sequence of each ASO is listed in Extended Data Table 1.

Mice. The first MeCP2-overexpressing mice were *MECP2*-TG mice on a FVB/N pure background[19]. These mice show normal locomotion in the open field and an increase in vertical activity, which was interpreted as less anxiety, but there was no difference in anxiety-like behaviour in the light–dark test. Mice on a pure FVB/N background, however, develop premature retinal degeneration, which can confound the interpretation of some behavioural tests[31]. To overcome issues related to a pure inbred strain, our laboratory characterized F_1 hybrid *MECP2*-TG mice (FVB/N × C57Bl/6 or FVB/N × 129S6/SvEv), and showed that these mice display several phenotypes as early as 7 weeks of age[20], including increased anxiety and a trend towards hypoactivity. Therefore, for both the genetic rescue and the ASO treatment experiments, we decided to continue using F_1 hybrid *MECP2*-TG mice.

For experiments related to the validation of the Flox;TG mouse model (Extended Data Figs 1 and 2), we generated F_1 hybrid animals by mating male *MECP2*-TG1 mice on a pure FVB/N background[19] to female mice heterozygous for the *Mecp2*^{lox} allele (Flox) (B6;129S4-*Mecp2*^{tm1Jae}/Mmucd obtained from MMRRC, and backcrossed to C57Bl/6J for more than 10 generations; see also scheme in Extended Data Fig. 1a). For experiments related to conditional rescue of *MECP2*-TG mice (Figs 1 and 2), we first mated Flox C57Bl/6 females with C57Bl/6 Cre-ER males (B6.Cg-Tg(UBC-cre/ERT2)1Ejb/J obtained from Jackson Laboratories). The F_1 Flox;Cre females were then mated to FVB/N *MECP2*-TG1 males to generate the F_1 hybrid, triple-transgenic Flox;TG;Cre male mice and their control littermates (Flox and Flox;TG) (see scheme in Fig. 1a). For studies related to *MECP2*-ASOs, we generated F_1 hybrid animals by mating FVB/N *MECP2*-TG1 females and wild-type 129S6/SvEv male mice (Taconic Farms). For the EEG experiment (Fig. 4b), we used *MECP2*-TG1 males on a pure FVB/N background.

We routinely used mouse littermates as controls for our experiments. Throughout the experiments, mice were maintained in a temperature-controlled, AALAS-certified level 3 facility on a 12 h light–dark cycle. Food and water were given ad libitum. All procedures to maintain and use these mice were approved by the Institutional Animal Care and Use Committee for Baylor College of Medicine. Animals were randomly selected using Excel software to generate a table of random numbers for all genetic and treatment studies. For all experiments, the individuals performing the behavioural and electrophysiological studies were blinded to the genotype or treatment.

Preparation of brain lysates and western blot. Brains were dissected and homogenized in cold lysis buffer (20 mM Tris-HCl, pH 8.0, 180 mM NaCl, 0.5% NP-40, 1 mM EDTA and Complete Protease Inhibitor, Roche). Lysates were rotated for 20 min at 4 °C. After centrifugation at 4 °C, the supernatant was mixed with NuPAGE sample buffer, heated for 5 min at 95 °C, and run on a NuPAGE 4–12% Bis-Tris gradient gel with MES SDS running buffer (NuPAGE). Proteins were transferred to a nitrocellulose membrane using NuPAGE Transfer Buffer for 1.5 h at 4 °C. The membrane was blocked for 1 h with 5% milk in TBS with 2% Tween-20 (TBST) followed by overnight incubation with primary antibody at 4 °C. After four 10-min washes with TBST, the membrane was incubated with secondary antibody for 1–2 h at room temperature. Horseradish peroxidase (HRP) was detected using SuperSignal West Dura kit, Thermo Scientific. Western blot images were acquired by ImageQuant LAS 4000 (GE Healthcare) and quantified by an ImageJ software package. Primary antibodies: rabbit antiserum raised against the amino terminus of MeCP2 (1:5,000; Zoghbi laboratory), mouse anti-GAPDH 6C5 (1:20,000; Advanced Immunochemicals, 2-RGM2). Secondary antibodies: goat anti-rabbit HRP (1:20,000; Bio-Rad), donkey anti-mouse HRP (1:20,000; Jackson ImmunoResearch Labs, 715-035-150).

Gene expression analysis by RT–qPCR. The subset of mice for RT–qPCR was selected randomly, using Excel software to generate a table of random numbers. No significant differences on behavioural measurements were found between the selected mice and the rest (per genotype group). Total RNA from mouse brain tissue was extracted using miRNeasy minikit (Qiagen), and 1 μg of total RNA was used to synthesize cDNA by Quantitect reverse transcription kit (Qiagen). For human lymphoblasts, 2 μg of total RNA was used to synthesize cDNA. RT–qPCR was performed in a CFX96 Real-Time System (Bio-Rad) using PerfeCTa SYBR Green Fast Mix (Quanta Biosciences). Sense and antisense primers were selected to be located on different exons, and the RNA was treated with DNase, to avoid false-positive results caused by DNA contamination. The specificity of the amplification products was verified by melting curve analysis. All RT–qPCR reactions were conducted in technical triplicates and the results were averaged for each sample, normalized to *Hprt* levels, and analysed using the comparative $\Delta\Delta C_t$ method. The following primers were used in the RT–qPCR reactions: *MECP2* (common to human and mouse): 5′-TATTTGATCAATCCCCAGGG-3′ (sense), 5′-CTCCCTCTCCCAGTTACCGT-3′ (antisense); *MECP2* (human-specific): 5′-GATGTGTATTTGATCAATCCC-3′ (sense), 5′-TTAGGGTCCAGGGATGTGTC-3′ (antisense); *Mecp2*-e1 (mouse-specific): 5′-AGGAGAGACTGGAGGAAAAGTC-3′ (sense), 5′-CTTAAACTTCAGTGGCTTGTCTCTG-3′ (antisense); *Mecp2*-e2 (mouse-specific): 5′-CTCACCAGTTCCTGCTTTGATGT-3′ (sense), 5′-CTTAAACTTCAGTGGCTTGTCTCTG-3′ (antisense); *MECP2*-e1 (human-specific): 5′-AGGAGAGACTGGAAGAAAAGTC-3′ (sense), 5′-CTTGAGGGGTTTGTCCTTGA-3′ (antisense); *MECP2*-e2 (human-specific): 5′-CTCACCAGTTCCTGCTTTGATGT-3′ (sense), 5′-CTTGAGGGGTTTGTCCTTGA-3′ (antisense); *Hprt* (mouse-specific): 5′-CGGGGGACATAAAAGTTATTG-3′ (sense), 5′-TGCATTGTTTTACCAGTGTCAA-3′ (antisense); *HPRT* (human-specific): 5′-GACCAGTCAACAGGGGACAT-3′ (sense), 5′-CCTGACCAAGGAAAGCAAAG-3′ (antisense); *Sst*: 5′-CCCAGACTCCGTCAGTTTCT-3′ (sense), 5′-GAAGTTCTTGCAGCCAGCTT-3′ (antisense); *Crf*: 5′-TACCAAGGGAGGAGAAGAGA-3′ (sense), 5′-GATCAGAACCGGCTGAGGT-3′ (antisense); *Npbwr1*: 5′-TCTCTTACTTCATCACCAGCC-3′ (sense), 5′-GCATAGAGGAAAGGGTTGAG-3′ (antisense); *Gamt*: 5′-GGATTATTGAGTGCAATGATGG-3′ (sense), 5′-TCAAGGGAACAACCTTATGTG-3′ (antisense); *Agrp*: 5′-TCAAGAAGACAACTGCAGAC-3′ (sense), 5′-TCTGTGGATCTAGCACCTC-3′ (antisense); *Rcor2*: 5′-ACCCGAAGTCGAACTAGTG-3′ (sense), 5′-CTAGTTCATCACTGTCTTCTTTG-3′ (antisense); *Prl2c2*: 5′-CATGAGCACCATGCTTCAG-3′ (sense), 5′-GCGAGCATCTTCATTGTCAG-3′ (antisense).

Immunofluorescence. Animals were anaesthetized with a mix of ketamine 37.6 mg ml^{-1}, xylazine 1.92 mg ml^{-1} and acepromazine 0.38 mg ml^{-1}, and transcardially perfused with 20 ml PBS followed by 100 ml of cold PBS-buffered 4% paraformaldehyde (PFA). The brains were removed and post-fixed overnight in 4% PFA. Next, brains were cryoprotected in 4% PFA with 30% sucrose at 4 °C for two additional days and embedded in Optimum Cutting Temperature (O.C.T., Tissue-Tek). Free-floating 40-μm brain sections were cut using a Leica CM3050 cryostat and collected in PBS. The sections were blocked for 1 h in 2% normal goat serum, 0.3% Triton X-100 in PBS at room temperature. Sections were then incubated overnight at 4 °C with either rabbit anti-MeCP2 antibody (1:1,000; Cell Signaling) or rabbit anti-ASO antibody (1:10,000; Isis Pharmaceuticals). The sections were washed three times for 10 min with PBS, and incubated for 3 h at room temperature with goat anti-rabbit antibody (1:500; Alexa Fluor 488, Invitrogen, A-11034). Sections were washed again three times for 10 min with PBS and mounted onto glass slides with Vectashield mounting medium with DAPI (Vector Laboratories).

Tamoxifen treatment. Tamoxifen (Sigma-Aldrich, T5648) was dissolved to 20 mg ml^{-1} in peanut oil, aliquotted and frozen at −20 °C until use. Peanut oil was also used as a vehicle. Tamoxifen or vehicle was injected intraperitoneally at a dose of 100 mg kg^{-1}, three alternative days a week for 4 weeks (as described in Fig. 1b).

Behavioural assays. All data acquisition and analyses were carried out by an individual blinded to the genotype and treatment. All behavioural studies were performed during the light period. Mice were habituated to the test room for 1 h before each test. At least one day was given between assays for the mice to recover. All the tests were performed as previously described[23] with few modifications.

Open field test. After habituation in the test room (150 lx, 60 dB white noise), mice were placed in the centre of an open arena (40 × 40 × 30 cm), and their behaviour was tracked by laser photobeam breaks for 30 min. General locomotor activity was automatically analysed using AccuScan Fusion software (Omnitech) by counting the number of times mice break the laser beams (activity counts). In addition, rearing activity, the time spent in the centre of the arena, entries to the centre and distance travelled were analysed. In this study, we found that *MECP2*-TG mice are hypoactive in the open field test. In contrast, in ref. 20, *MECP2*-TG mice show a non-significant trend towards hypoactivity. This difference might be the result of our study assessing locomotor activity by measuring activity counts, and in ref. 20 by measuring the distance travelled, which is calculated by the software from the activity counts.

Elevated plus maze test. After habituation in the test room (700 lx, 60 dB white noise), mice were placed in the centre part of the maze facing one of the two open arms. Mouse behaviour was video-tracked for 10 min, and the time mice spent in the open arms and the entries to the open arms, as well as the distance travelled in the open arms, were recorded and analysed using ANY-maze system (Stoelting).

Accelerating rotarod test. After habituation in the test room (700 lx, 60 dB white noise), motor coordination was measured using an accelerating rotarod apparatus (Ugo Basile). Mice were tested for two consecutive days, four trials each, with an interval of 60 min between trials to rest. Each trial lasted for a maximum of 10 min,

and the rod accelerated from 4 to 40 r.p.m. in the first 5 min. The time that it took for each mouse to fall from the rod (latency to fall) was recorded.

Three-chamber test. The three-chamber apparatus consists of a clear Plexiglas box ($24.75 \times 16.75 \times 8.75$) with removable partitions that separate the box into three chambers. In both the left and right chambers a cylindrical wire cup was placed with the open side down. Age- and gender-matched C57Bl/6 mice were used as novel partners. Two days before the test, the novel partner mice were habituated to the wire cups (3 inches diameter by 4 inches in height) for 1 h per day. After habituation in the test room (700 lx, 60 dB white noise), mice were placed in the central chamber and allowed to explore the three chambers for 10 min (habituation phase). Next, a novel partner mouse was placed into a wire cup in either the left or the right chamber. An inanimate object was placed as control in the wire cup of the opposite chamber. The location of the novel mouse was randomized between left and right chambers across subjects to control for side preference. The mouse tested was allowed to explore again for an additional 10 min. The time spent investigating the novel partner (defined by rearing, sniffing or pawing at the wire cup) and the time spent investigating the inanimate object were measured manually.

RNA-seq. Mice were euthanized under anaesthesia and the hippocampi were quickly dissected over ice. Total RNA of 30 hippocampal samples (three biological replicates of each genotype from three experiments) was extracted using miRNeasy minikit (Qiagen), following the manufacturer's instructions. Isolated RNA was eluted in RNase-free water and submitted to the Genomic and RNA Profiling Core at Baylor College of Medicine. Sample quality checks using the NanoDrop spectrophotometer and Agilent Bioanalyzer 2100 were conducted. Then Illumina TruSeq RNA library preparation protocol was used as follows. A double-stranded DNA library was created using 250 ng of total RNA (measured by picogreen), preparing the fragments for hybridization onto a flow-cell. First, cDNA was created using the fragmented 3′ poly(A) selected portion of total RNA and random primers. During second-strand synthesis, dTTP is replaced with dUTP, which quenches the second strand during amplification, thereby achieving strand specificity. Libraries were created from the cDNA by first blunt-ending the fragments, attaching an adenosine to the 3′-end and finally ligating unique adapters to the ends. The ligated products were then amplified using 15 cycles of PCR. The resulting libraries were quantified using the NanoDrop spectrophotometer and fragment size assessed with the Agilent Bioanalyzer. A qPCR quantification was performed on the libraries to determine the concentration of adaptor ligated fragments using Applied Biosystems ViiA7 Real-Time PCR System and a KAPA Library Quant Kit. Using the concentration from the ViiA7 qPCR machine, 21 pM of library was loaded onto a flow-cell and amplified by bridge amplification using the Illumina cBot machine. A paired-end 100 cycle run was used to sequence the flow-cell on a HiSeq Sequencing System.

RNA-seq data pre-processing and analysis. For each sample, about 10 million 100-base-pair pair-end reads were generated. Raw reads were first groomed by removing adapters from both the 3′- and 5′-ends before mapping to the reference genome. Then, trimmed reads were aligned to the *Mus musculus* genome (UCSC *mm10*; the gene model for the mapping was obtained from http://ccb.jhu.edu/software/tophat/igenomes.shtml) using TopHat v2.0.9 (ref. 32) with default parameters (-r 200 -p 5). The mappability for all 30 samples was above 85%. To prepare the aligned sequence reads into expression level for differential gene analysis, we used the free Python program HTSeq[33]. The htseq-count function of HTSeq allowed us to accumulate the number of aligned reads that fall under the exons of the gene (union of all the exons of the gene). These read counts are analogous to the expression level of the gene. Using the obtained read counts, differential gene analyses were carried out using the DESeq package and glm.nb function in the R environment. DESeq includes functions for us to normalize the read counts of multiple samples across several genotypes by the use of the negative binomial distribution and a shrinkage estimator for the distribution's variance[34]. glm.nb allows us to fit a negative binomial regression model to test the gene changes between genotypes. Specifically, for data from the ASO experiment, each gene was tested to check whether its expression levels in wild-type and TG-ASO mice differed from that in transgenic mice. Similarly, for data from the genetic rescue experiment, expression levels in Flox and Flox;TG;Cre-TMX were tested for differences from the expression in Flox;TG;Cre-vehicle and Flox;TG. The statistical significance of the observed changes was reported by the false discovery rate, which is the *P* value adjusted for multiple testing with the Benjamini–Hochberg procedure. A gene was considered significantly different between genotypes if it fell under a false discovery rate of 10% and changed in a coherent direction.

Sample clustering. To assess the similarity of expression patterns between samples of different genotypes, we carried out unbiased clustering: expressions of the identified significantly changed genes were clustered by sample based on Euclidean distance on average linkage and by genes based on Euclidean distance on complete linkage. Heat maps were then used to plot the clustered gene expressions for visual inspection. The plotted expressions (*Z*-scores) for each gene were the expressions normalized at the gene level to have an average of zero and a standard deviation of one.

Hippocampal slice preparation. Mice were deeply anaesthetized with isofluorane, followed by decapitation. The brain was removed into oxygenated and ice-cold cutting solution (CS) containing (in mM): 110 sucrose, 60 NaCl, 3 KCl, 1.25 NaH_2PO_4, 28 $NaHCO_3$, 0.5 $CaCl_2$, 7 $MgCl_2$ and 5 glucose, and the caudal portion of the forebrain containing the hippocampus and entorhinal cortex was isolated by razor blade cuts. Transverse slices (400 μm) were prepared with a Vibratome (Vibratome). Cortical tissue was then removed and hippocampal slices were equilibrated in a mixture of 50% CS and 50% artificial cerebrospinal fluid (ACSF) containing (in mM): 125 NaCl, 1.25 NaH_2PO_4, 2.5 KCl, 25 $NaHCO_3$, 2 $CaCl_2$, 1 $MgCl_2$ and 15 glucose, at room temperature for 10–20 min before transfer to the recording chamber.

Slice electrophysiology. All data acquisition and analyses were carried out blinded to the genotype and treatment. Electrophysiology was performed in an interface chamber (Fine Science Tools). Oxygenated ACSF (95%/5% O_2/CO_2, 31 °C) was perfused into the recording chamber at the rate of 1.5 ml min^{-1}. Electrophysiological traces were digitized and stored using a Digidata 1320A and Clampex software (Axon Instruments). fEPSPs were recorded in the stratum radiatum with an ACSF-filled glass recording electrode (1–3 MΩ). The relationship between fibre volley amplitude and fEPSP slope over various stimulus intensities was used to assess baseline synaptic transmission. All subsequent experimental stimuli were set to an intensity that evoked a 30–40% of the maximal fEPSP slope. Slices that did not exhibit stable fEPSP slopes during the first 20 min of recording were excluded from the analysis. Paired-pulse facilitation was measured at varying interstimulus intervals (20, 50, 100, 200 and 300 ms). LTP was induced by two trains of high-frequency stimulation (100 Hz for 1 s) with a 20-s intertrain interval. Synaptic efficacy was monitored for 20 min before and 70 min after LTP induction by recording fEPSPs every 20 s (three traces were averaged over succeeding 1-min intervals). For the quantification of the last 10 min of the LTP recording (Fig. 2i), slices were averaged per mouse, and statistical analysis was done on the animals ($n = 6$–7 mice, 15–20 slices, two-tailed *t*-test).

Intracerebral injection of ASO. Mice were anaesthetized with isoflurane and placed on a computer-guided stereotaxic instrument (Angle Two Stereotaxic Instrument, Leica Microsystems) that is fully integrated with the Franklin and Paxinos[35] mouse brain atlas through a control panel. Anaesthesia (isoflurane 3%) was continuously delivered via a small face mask. Ketoprofen 5 mg kg^{-1} was administered subcutaneously at the initiation of surgery. After sterilizing the surgical site with betadine and 70% alcohol, a midline incision was made over the skull and a small hole was drilled through the skull above the right lateral ventricle. A total of 500 μg *MECP2*-ASO or saline was delivered using a Hamilton syringe connected to a motorized nanoinjector at 0.3 μl min^{-1}. The coordinates used relative to bregma were: anteroposterior (AP) = −0.2 mm, medial lateral (ML) = 1 mm, dorsal ventral (DV) = −3 mm, based on a calibration study indicating these coordinates as leading to the right ventricle in our mice. To allow diffusion of the solution into the brain, the needle was left for 5 min on the site of injection. The incision was manually closed with suture. Carprofen-containing food pellets were provided for 5 days after the surgery. Two weeks after the surgery,the animals were euthanized and their brains were dissected for RNA and protein analysis.

Surgical implantation of cannula and osmotic pumps. Two days before surgery, a micro-osmotic pump (Alzet model 1004, Durect) was filled with 500 μg *MECP2*-ASO or control-ASO dissolved in 100 μl saline. The pump was then connected through a plastic catheter to a cannula (Alzet Brain Infusion Kit 3, Durect) (see Extended Data Fig. 6a). The pump was designed to deliver the drug at a rate of 0.11 μl h^{-1} for 28 days. The cannula plus pump assembly was primed in sterile saline for 2 days at 37 °C. Mice were anaesthetized with isoflurane and placed on a computer-guided stereotaxic instrument (Angle Two Stereotaxic Instrument, Leica Microsystems). Anaesthesia (isoflurane 3%) was continuously maintained via a small face mask. Ketoprofen 5 mg kg^{-1} was administered subcutaneously at the initiation of the surgery. After sterilizing the surgical site with betadine and 70% alcohol, a midline incision was made over the skull and a subcutaneous pocket was generated on the back of the animal. Next, the pump was inserted into the pocket and the cannula was stereotactically implanted to deliver the drug in the right ventricle using the following coordinates: AP = −0.2 mm, ML = 1 mm, DV = −3 mm. The incision was sutured shut. Carprofen-containing food pellets were provided for 5 days after the surgery. The pump was disconnected and removed 28 days after the initiation of treatment. Two additional weeks were given to the animals to recover before any behavioural testing.

EEG monitoring. Mice were anaesthetized with isoflurane and mounted in a stereotaxic frame. Under aseptic conditions, each mouse was surgically implanted with three recording electrodes (Teflon-coated silver wire, 125 μm in diameter) aimed at the subdural space of left frontal cortex, left parietal cortex and right parietal cortex. The reference electrode was then positioned in the occipital region of the skull.

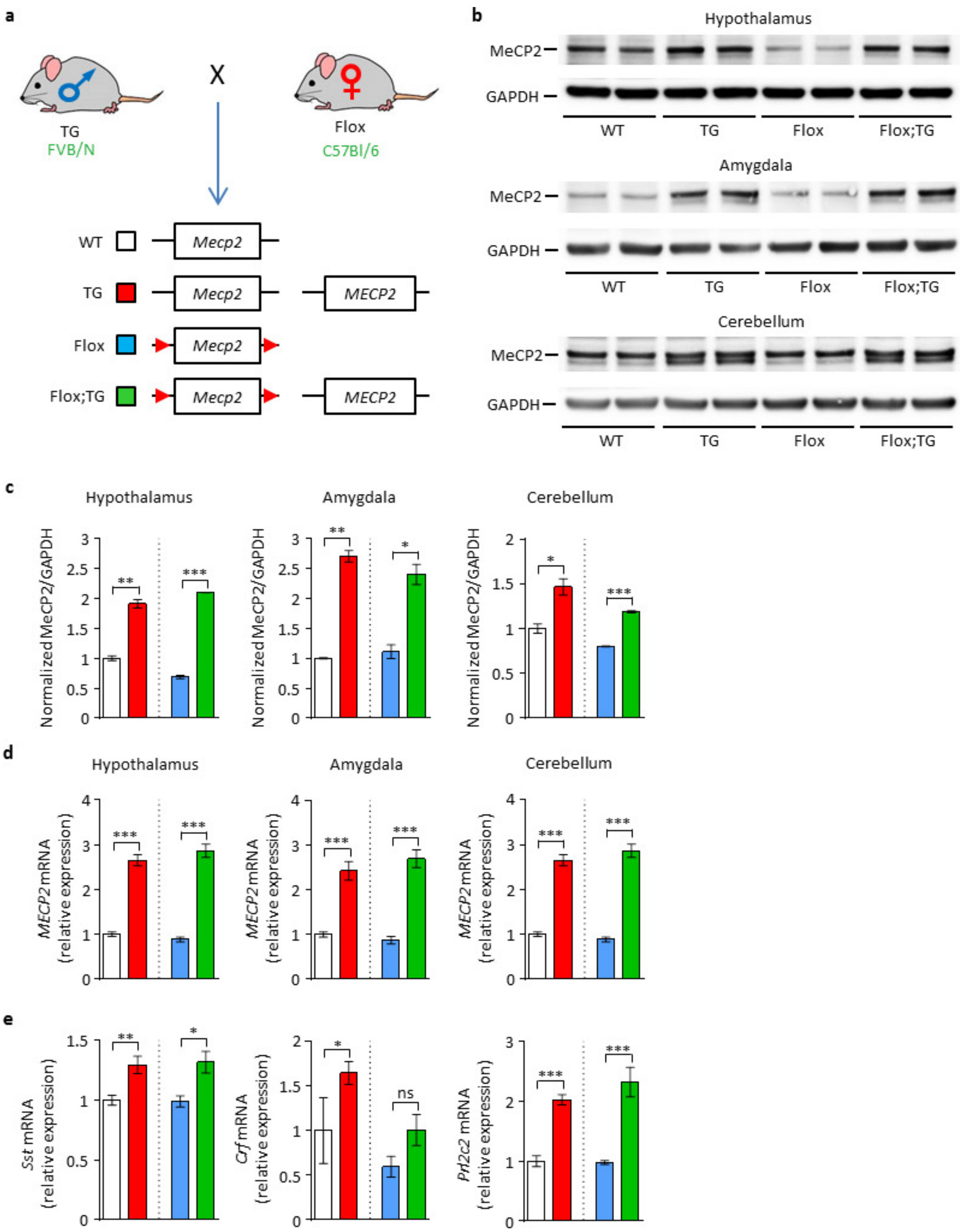
All electrode wires were attached to a miniature connector (Harwin Connector). After 3–5 days of post-surgical recovery, cortical EEG activity (filtered between 0.5 and 5 kHz, sampled at 2 kHz) and behaviour were simultaneously recorded in freely moving mice for 2 h per day over 3–5 days[36].

EEG data analysis. All the EEG recordings were qualitatively and manually analysed by experimenters blinded to the mouse genotype and treatment. Electrographic seizure activities were visually identified and matched with the behavioural seizure, if applicable. Other abnormal epileptiform spikes were also identified visually[37].

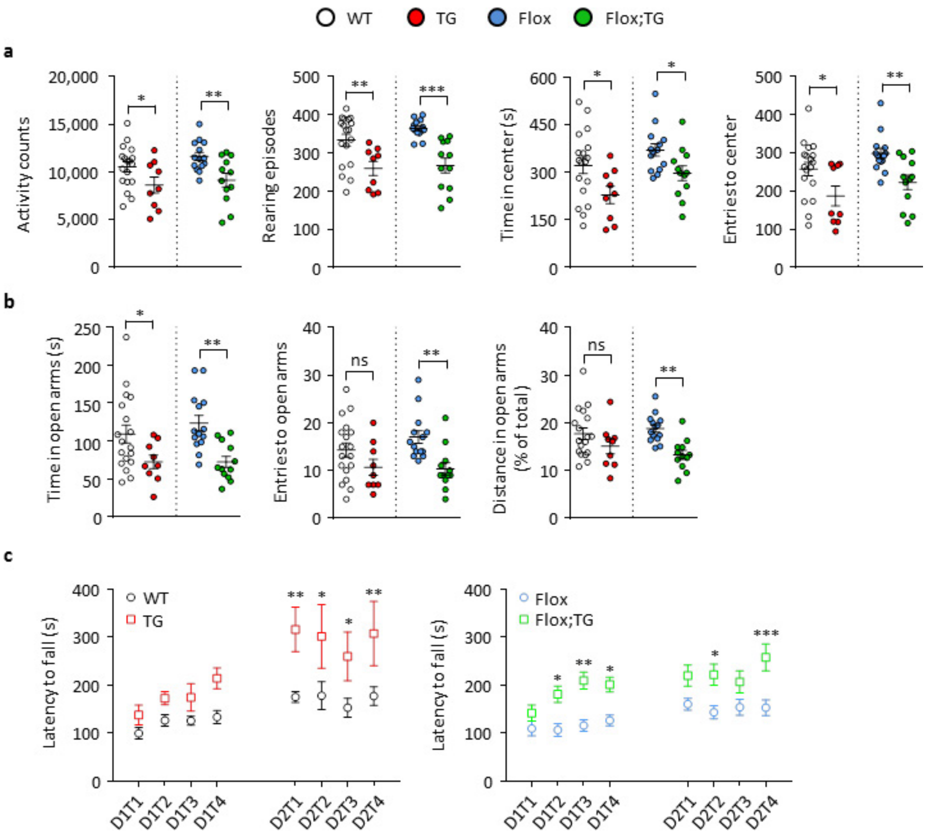
Lymphoblasts culture and transfection with ASOs. Following informed consent, approved by the Institutional Review Board for Human Subject Research at Baylor College of Medicine (H-18122), a venous blood sample was provided by five individuals affected with *MECP2* duplication syndrome and five age-matched controls to establish immortalized B-lymphoblastoid cell lines, following standard procedures. Human B-lymphoblastoid cells were cultured in suspension in RPMI 1640 medium with L-glutamine, penicillin–streptomycin and 10% (v/v) FBS. A day before transfection, cells were seeded in 6-wells plates at a density of 1×10^6 cells in a total volume of 2 ml complete medium. Transfection mixture was prepared by combining 20 μl ASO (at the desire concentration), 4 μl transfection reagent (TurboFect, R0531, Thermo Scientific) and 180 μl serum-free RPMI medium. The mix was incubated at room temperature for 15 min before adding to the cells. RNA was extracted from lymphoblasts 48 h after transfection. Lymphoblastoid cells from the age-matched control donors and the non-treated *MECP2* duplication cells were incubated with 4.8 μM control-ASO.

Statistical analysis. Statistical significance was analysed using GraphPad Prism. The number of animals used (n), and the specific statistical tests used are indicated for each experiment in the figure legends. Sample size in behavioural studies was based on previous reports using transgenic mice with the same background. Mice were randomly assigned to vehicle or treatment groups using Excel software to generate a table of random numbers, and the experimenter was always blinded to the treatment. For behavioural assays, all population values appear normally distributed. Equal variances were never assumed, and the Geisser–Greenhouse correction for sphericity was always applied when using ANOVA.

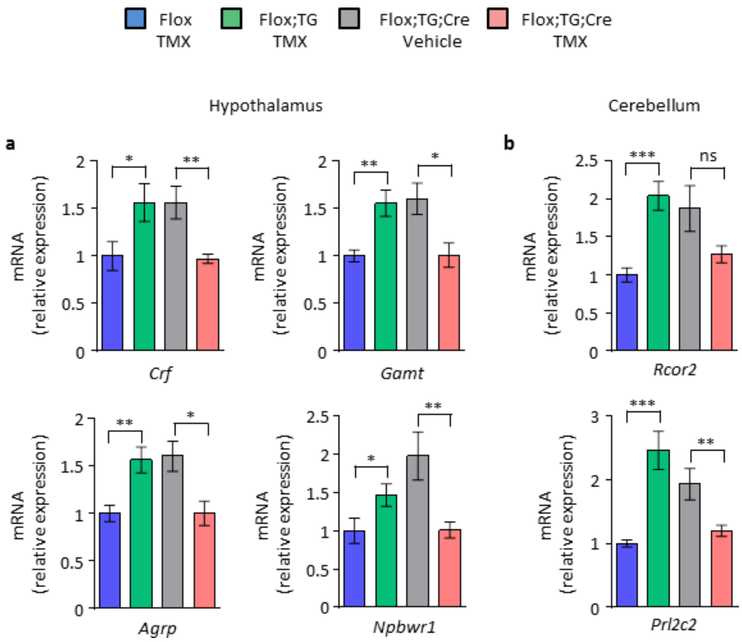
31. Cook, M. N., Williams, R. W. & Flaherty, L. Anxiety-related behaviors in the elevated zero-maze are affected by genetic factors and retinal degeneration. *Behav. Neurosci.* **115,** 468–476 (2001).
32. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).
33. Anders, S., Pyl, P. T. & Huber, W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).
34. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).
35. Paxinos, G. & Franklin, K. B. *The Mouse Brain in Stereotaxic Coordinates* (Academic, 2001).
36. Han, K. *et al.* SHANK3 overexpression causes manic-like behaviour with unique pharmacogenetic properties. *Nature* **503,** 72–77 (2013).
37. Roberson, E. D. *et al.* Amyloid-β/Fyn-induced synaptic, network, and cognitive impairments depend on tau levels in multiple mouse models of Alzheimer's disease. *J. Neurosci.* **31,** 700–711 (2011).



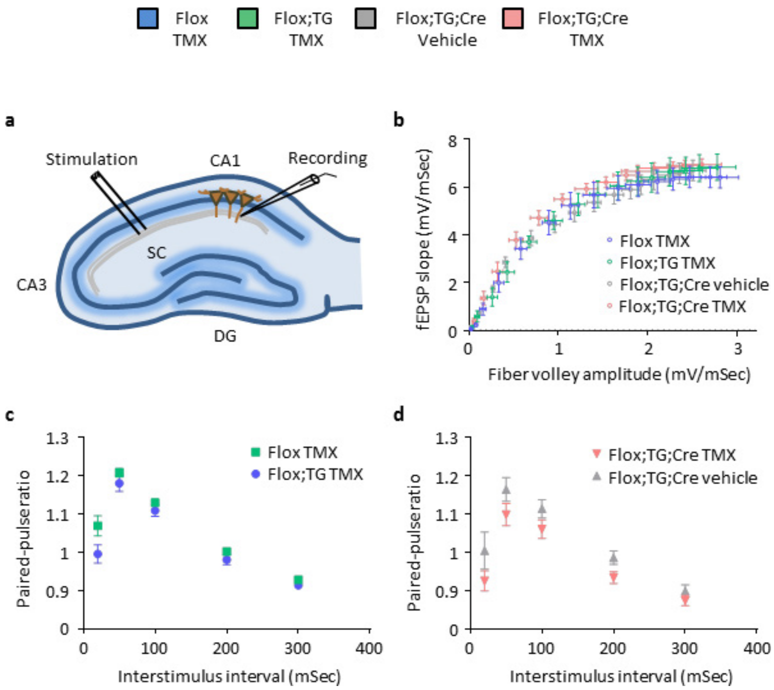
Extended Data Figure 1 | Mice that overexpress a human *MECP2* transgene over a floxed *Mecp2* allele resemble the classic *MECP2*-TG mice at the molecular level. a, Breeding strategy to generate mice that have a wild-type *MECP2* allele and a *Mecp2* allele flanked by *loxP* sequences. **b**, Western blot of MeCP2 in hypothalamus, amygdala and cerebellum. GAPDH was used as the internal control (for gel source data, see Supplementary Fig. 3). **c**, Densitometric analysis of western blots in **b**. Flox;TG mice overexpress MeCP2 at levels similar to transgenic mice. It is noteworthy that in the hypothalamus, Flox mice were 30% hypomorphic ($n = 2$ mice per group; two-tailed t-test) when compared to wild-type mice, but not in the other regions. **d**, RT–qPCR analysis in hypothalamus, amygdala and cerebellum, using primers common to mouse and human *MECP2*. Flox;TG mice overexpressed the *MECP2* transcript at levels similar to transgenic mice. *Hprt1* was used as the internal control ($n = 5$ mice per group; two-tailed t-test). **e**, RT–qPCR analysis of three selected genes know to be altered by MeCP2 overexpression. Flox;TG mice overexpressed the *Sst*, *Crf* and *Prl2c2* transcripts at levels similar to those of transgenic mice. *Hprt1* was used as the internal control ($n = 4$ for wild-type group; $n = 6$ for transgenic group; $n = 5$ for Flox and Flox;TG groups; two-tailed t-test). Data are mean ± s.e.m. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.



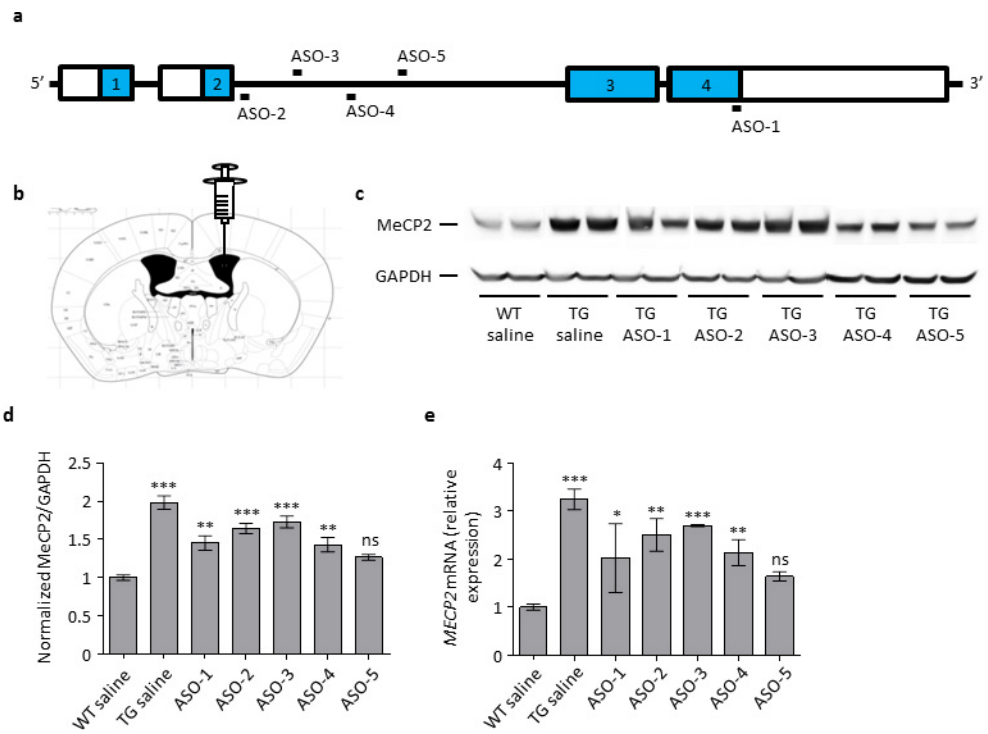
Extended Data Figure 2 | Mice that overexpress a human *MECP2* transgene over a floxed *Mecp2* allele show characteristic behavioural deficits. a, Flox;TG mice displayed hypoactivity and anxiety in the open field test similar to transgenic mice. **b**, Flox;TG mice showed heightened anxiety-like behaviour in the elevated plus maze test. **c**, Flox;TG mice showed enhanced motor learning in the rotarod test similar to transgenic mice ($n = 18$ for wild-type group; $n = 9$ for transgenic group; $n = 14$ for Flox group; $n = 12$ for Flox;TG group (for all behavioural tests)). Data were analysed by one-way ANOVA, with the exception of the rotarod test that was analysed by two-way ANOVA repeated measures followed by Tukey HSD post hoc correction for multiple comparisons. Data are mean ± s.e.m. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.



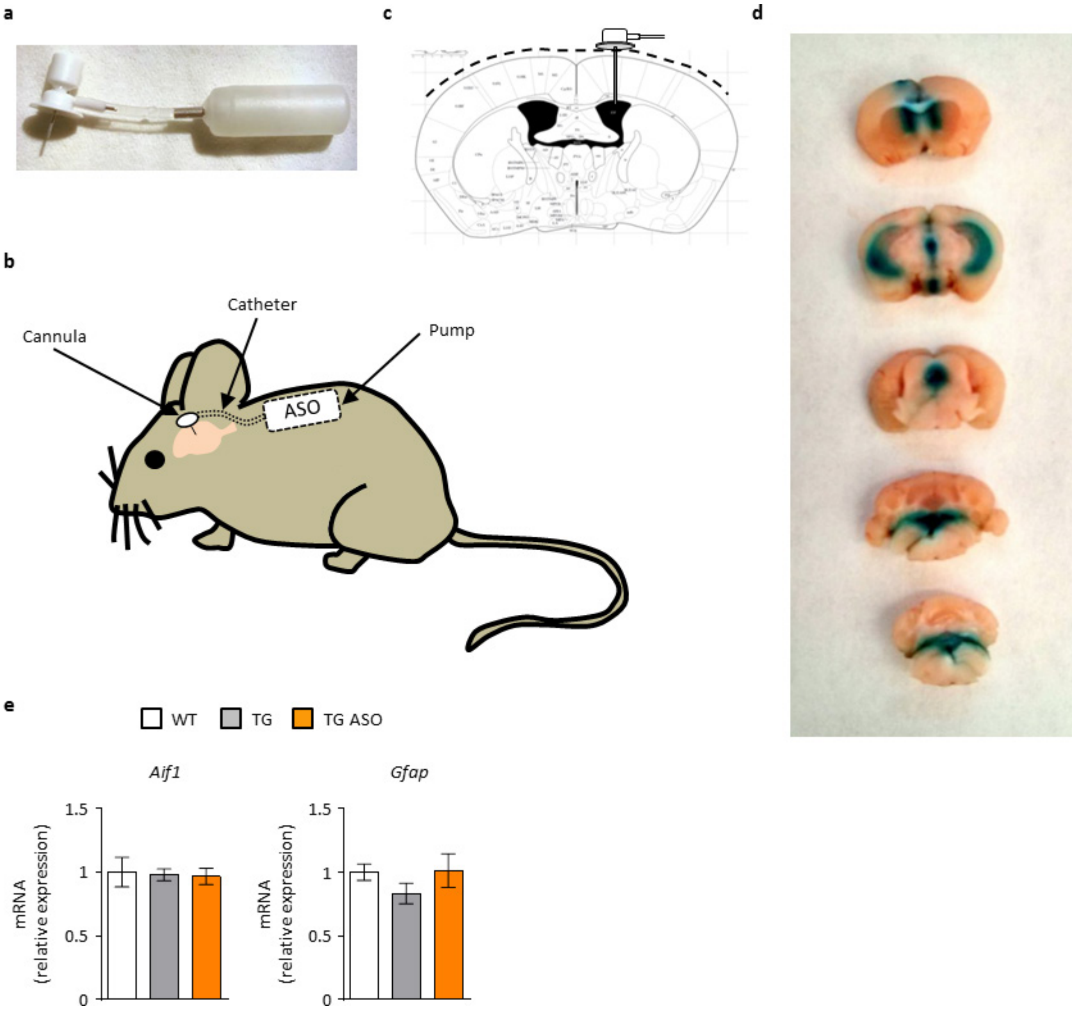
Extended Data Figure 3 | Correction of gene expression after genetic rescue. a, **b**, RT–qPCR from the hypothalamus (**a**) and cerebellum (**b**) shows correction of altered expression of selected genes after normalization of MeCP2 levels in Flox;TG;Cre-TMX mice ($n = 6$ for Flox group; $n = 7$ for Flox;TG group; $n = 7$ for Flox;TG;Cre-vehicle group; $n = 8$ for Flox;TG;Cre-TMX group; two-tailed t-test). Data are mean ± s.e.m. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.



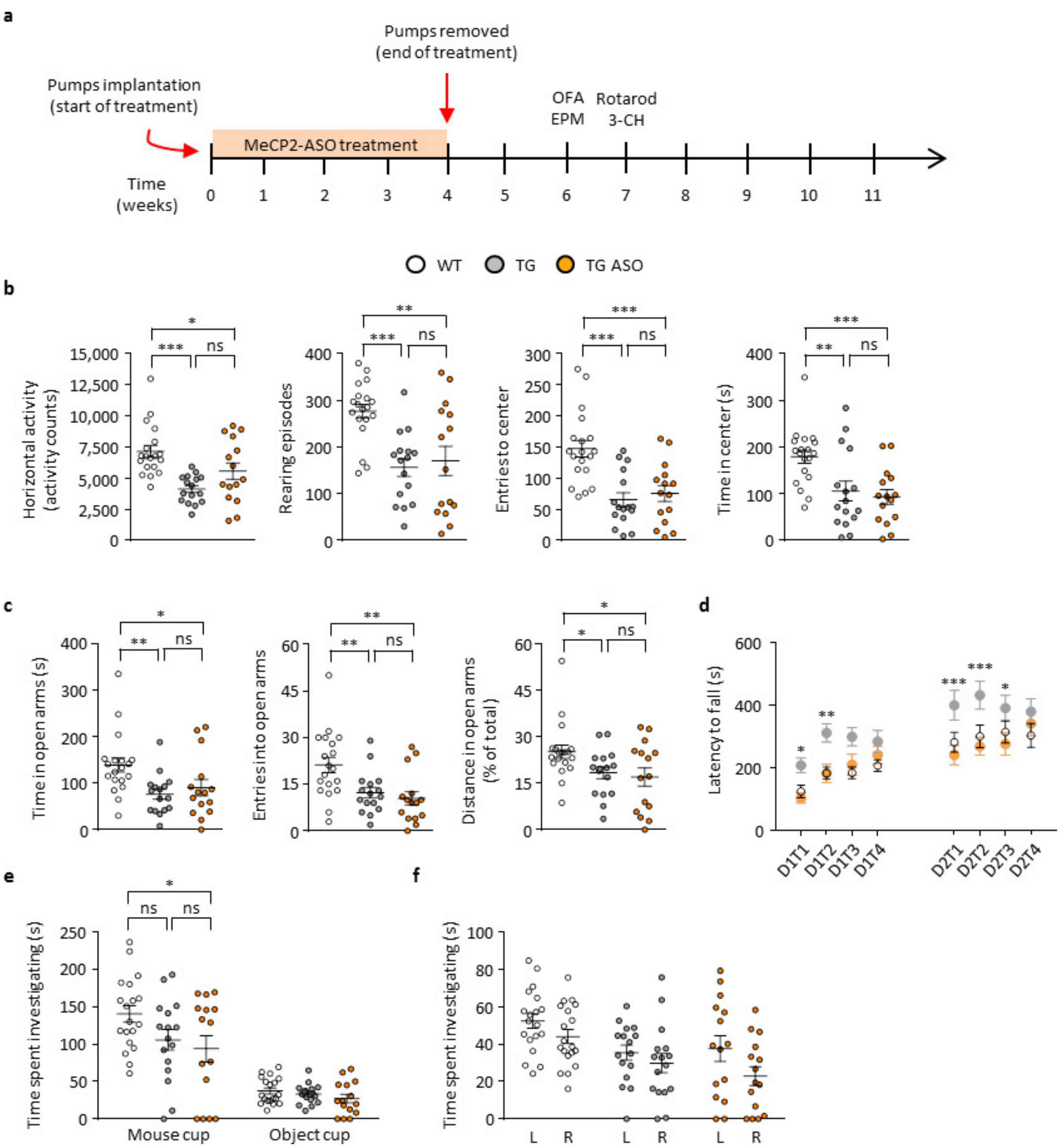
Extended Data Figure 4 | Hippocampal basal synaptic transmission and short-term synaptic plasticity are normal in *MECP2*-TG mice. a, Long-term potentiation was induced by applying high-frequency stimulation to the Schaffer collateral axons, and field excitatory postsynaptic potentials (fEPSPs) were recorded at the Schaffer collateral-CA1 synapses of the hippocampus (stratum radiatum). **b**, MeCP2 overexpression did not affect Schaffer collateral basal synaptic transmission, as determined by the correlation between the slopes of the evoked fEPSPs and the amplitudes of the fibre volleys. **c**, **d**, MeCP2 overexpression did not affect short-term synaptic plasticity, as determined by paired-pulse facilitation ($n = 7$ for Flox group; $n = 6$ for Flox;TG group; $n = 7$ for Flox;TG;Cre-vehicle group; $n = 7$ for Flox;TG;Cre-TMX group). Data are mean $\pm$ s.e.m.



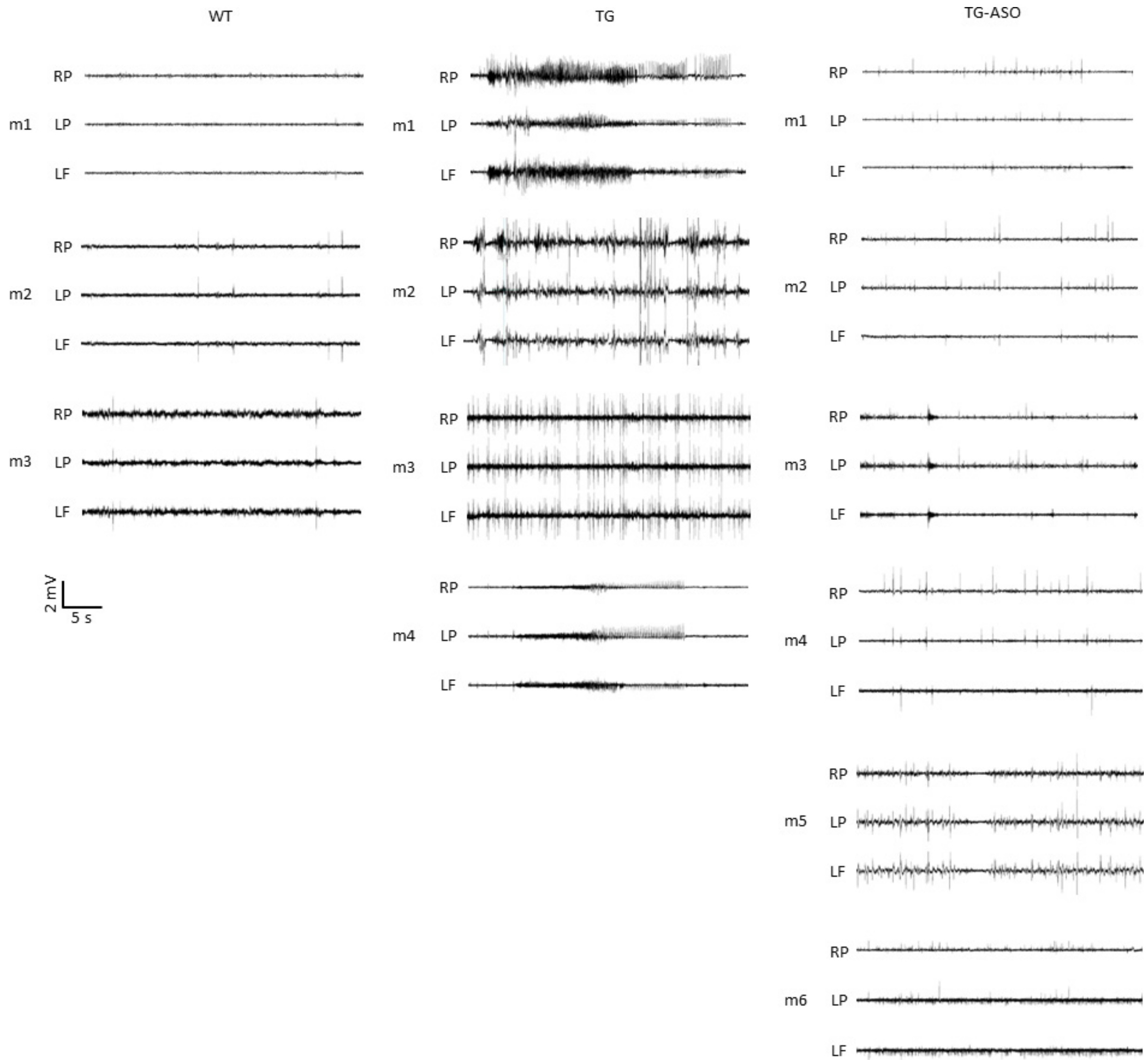
Extended Data Figure 5 | ASO-5 reduces *MECP2* mRNA and protein levels. **a**, Location of ASOs targeting sequences on the *MECP2* pre-mRNA. Boxes represent exons, and lines denote introns. Boxes in blue denote the translatable regions. **b**, Section schemata from the Paxinos and Franklin[35] mouse brain atlas showing site of stereotactic injection. **c**, Western blot (**c**) and densitometric analysis (**d**) of MeCP2 from cortical samples of mice treated with saline or the indicated ASOs 2 weeks after single bolus stereotactic injection of 500 μg ASO in the right ventricle of the brain. ASO-5 was found to be the most effective. GAPDH was used as an internal control ($n = 3$, one-way ANOVA followed by Tukey HSD post hoc correction for multiple comparisons; for gel source data, see Supplementary Fig. 4). **e**, RT–qPCR analysis of *MECP2* mRNA from cortical samples of mice treated with saline or the indicated ASOs, 2 weeks after single bolus stereotaxic injection of 500 μg ASO in the right ventricle of the brain. ASO-5 was found to be the most effective. The *MECP2* primers are common to the mouse and human alleles. *Hprt1* was used as an internal control ($n = 3$, one-way ANOVA followed by Tukey HSD post hoc correction for multiple comparisons). Data are mean ± s.e.m. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.



Extended Data Figure 6 | Osmotic micro-pumps for gradual intracerebroventricular infusion of ASO. a, Micro-osmotic pump connected to a cannula through a plastic catheter. **b**, The micro-osmotic pump was implanted in a subcutaneous pocket and the cannula stereotaxically positioned to deliver the ASO into the right ventricle. **c**, Section schemata from the Paxinos and Franklin[35] mouse brain atlas showing site of cannula implantation. **d**, Injection of 1 µl of dye through the catheter to show that the dye reaches the whole ventricular system, confirming the correct positioning of the tip of the cannula in the right ventricle. **e**, RT–qPCR for *Aif1* (a marker of activated microglia) and *Gfap* (a marker of astrocytes) immediately after the end of 4 weeks of gradual ASO treatment ($n = 4$ for wild-type group; $n = 5$ for transgenic group; $n = 4$ for TG-ASO group). Data are mean ± s.e.m.



Extended Data Figure 7 | Behavioural phenotype 6–7 weeks after the initiation of ASO treatment. a, Timeline of ASO treatment and behavioural tests. **b**, Two weeks after cessation of ASO treatment, *MECP2*-TG mice showed no rescue of hypoactivity, rearing behaviour or anxiety parameters in the open field test. **c**, No rescue was evident in any of the parameters of the elevated plus maze test at this early post-treatment stage. **d**, ASO treatment normalized performance in the rotarod test in *MECP2*-TG mice (asterisks indicate significance between *MECP2*-TG ASO and control-ASO groups). **e**, The impaired social behaviour in the 3-chamber test was not normalized in the ASO-treated group. No significant difference was found between any of the groups in the time spent investigating the inanimate object. **f**, No preference for the cups placed in the left or right chambers was found in the habituation phase of the 3-chamber test ($n = 19$ for wild-type group; $n = 16$ for transgenic group; $n = 15$ for TG-ASO group (for all behavioural tests)). Data were analysed by one-way ANOVA followed by Fisher's LSD post hoc test, with the exception of the rotarod test that was analysed by two-way ANOVA repeated measures followed by Tukey HSD post hoc correction for multiple comparisons. Data are mean ± s.e.m. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.



Extended Data Figure 8 | Cortical EEG recording. Representative EEG traces of all mice recorded after gradual *MECP2*-ASO or control-ASO treatment.

Extended Data Table 1 | ASO sequences

ASO	Sequence	Length (bp)	Molecular weight (Da)	Chemistry
ASO-1	AACTCTCTCGGTCACGGGCG	20	7141.88	MOE gapmer
ASO-2	CACACTGACCTTTCAGGGCT	20	7100.87	MOE gapmer
ASO-3	GATCACTGGAACACAATGGT	20	7155.87	MOE gapmer
ASO-4	CGTGCCATGGAAGTCCTTCC	20	7116.87	MOE gapmer
ASO-5	GGTTTTTCTCCTTTATTATC	20	7035.77	MOE gapmer
Control-ASO	GTTTTCAAATACACCTTCAT	20	7045.82	MOE gapmer

The sequence, length, molecular mass and type of chemical modification of the different *MECP2* ASOs tested and the control ASO.

Extended Data Table 2 | Demographics of *MECP2* duplication syndrome patients

Subject ID	Date of birth	Date of collection	Age at collection (years)	Gender	*MECP2* expression fold change (relative to average control)
Control-1	07.10.1987	10.15.1992	5	Male	1.26
Control-2	08.22.1998	07.08.1999	1	Male	0.87
Control-3	07.13.2006	02.15.2013	6.7	Male	0.62
Control-4	11.17.2006	11.02.2012	6	Male	0.80
Control-5	06.23.2006	06.14.2013	7	Male	1.43
Duplicaion-1	04.11.2007	11.13.2008	1.5	Male	2.51
Duplication-2	11.07.2002	12.06.2007	5	Male	2.28
Duplication-3	10.25.2003	04.14.2008	4.5	Male	2.42
Duplication-4	02.02.1999	07.26.2006	7.5	Male	1.99
Duplication-5	11.01.2002	06.17.2008	5.5	Male	2.00

The demographics of the *MECP2* duplication patients and healthy donors that provided blood samples to establish immortalized B-lymphoblastoid cell lines. The last column on the right describes the individual *MECP2* mRNA expression levels in lymphoblasts, relative to the average of the healthy donors.

LETTER

doi:10.1038/nature15715

Therapeutic antibodies reveal Notch control of transdifferentiation in the adult lung

Daniel Lafkas[1], Amy Shelton[1], Cecilia Chiu[2], Gladys de Leon Boenig[3], Yongmei Chen[2], Scott S. Stawicki[2], Christian Siltanen[1]†, Mike Reichelt[4], Meijuan Zhou[5], Xiumin Wu[5], Jeffrey Eastham-Anderson[4], Heather Moore[6], Meron Roose-Girma[7], Yvonne Chinn[8], Julie Q. Hang[8], Søren Warming[7], Jackson Egen[6]†, Wyne P. Lee[5], Cary Austin[4], Yan Wu[2], Jian Payandeh[3], John B. Lowe[4] & Christian W. Siebel[1]

Prevailing dogma holds that cell–cell communication through Notch ligands and receptors determines binary cell fate decisions during progenitor cell divisions, with differentiated lineages remaining fixed[1]. Mucociliary clearance[2,3] in mammalian respiratory airways depends on secretory cells (club and goblet) and ciliated cells to produce and transport mucus. During development or repair, the closely related Jagged ligands (JAG1 and JAG2) induce Notch signalling to determine the fate of these lineages as they descend from a common proliferating progenitor[4–8]. In contrast to such situations in which cell fate decisions are made in rapidly dividing populations[9,10], cells of the homeostatic adult airway epithelium are long-lived[11–13], and little is known about the role of active Notch signalling under such conditions. To disrupt Jagged signalling acutely in adult mammals, here we generate antibody antagonists that selectively target each Jagged paralogue, and determine a crystal structure that explains selectivity. We show that acute Jagged blockade induces a rapid and near-complete loss of club cells, with a concomitant gain in ciliated cells, under homeostatic conditions without increased cell death or division. Fate analyses demonstrate a direct conversion of club cells to ciliated cells without proliferation, meeting a conservative definition of direct transdifferentiation[14]. Jagged inhibition also reversed goblet cell metaplasia in a preclinical asthma model, providing a therapeutic foundation[15]. Our discovery that Jagged antagonism relieves a blockade of cell-to-cell conversion unveils unexpected plasticity, and establishes a model for Notch regulation of transdifferentiation.

Using phage display to generate synthetic therapeutic antibodies targeting JAG1 or JAG2, we identified a selective and potent inhibitor of each ligand that also cross-reacted with the human and mouse orthologues: anti-JAG1 blocking antibody version 70 (anti-JAG1.b70) and anti-JAG2 blocking antibody version 33 (anti-JAG2.b33). Anti-JAG1.b70 bound to purified JAG1 with high affinity but not to JAG2 or the other canonical Notch ligands Delta-like1 (DLL1) or DLL4; conversely, anti-JAG2.b33 showed high-affinity binding only to JAG2, although very weak binding (3,500-fold lower affinity) was detected to JAG1 (Fig. 1a and Extended Data Figs 1 and 2a). Each antibody potently inhibited signalling induced only by the cognate ligand (Fig. 1b, c and Extended Data Fig. 2c). Both antibodies inhibited signalling through NOTCH1, NOTCH2 and NOTCH3 (data not shown), demonstrating that inhibition occurs irrespective of the particular receptor.

To understand the molecular basis of anti-JAG1 antagonism and selectivity, we determined a high-resolution crystal structure of the antibody Fab fragment bound to human JAG1. The epitope lies at the junction of the Delta-Serrate-Lag (DSL) and epidermal growth factor-like-1 (EGF1) domains (Fig. 1d, e and Extended Data Figs 2b and 3). Numerous epitope residues differ between the JAG orthologues. Notably, heavy-chain residue His35 forms an ionic bridge with JAG1 residue Asp204 (Fig. 1d, e and Extended Data Fig. 3b, c), an apparent anchor that cannot be formed with the equivalent JAG2 residue Asn204. Sequence alignments and structural modelling provide clear explanations for anti-JAG1.b70 cross-reactivity to human and mouse JAG1 (Extended Data Fig. 3a, c). The NOTCH1–DLL4 crystal structure[16] (Protein Data Bank (PDB) accession 4XL1) and models of the NOTCH2–JAG1 interaction[17] indicate that the DSL–EGF1 face contains residues key to NOTCH receptor binding, pointing to steric hindrance of receptor binding as the mechanism for antibody blocking (Extended Data Fig. 3d, e).

To determine the effects of JAG inhibition *in vivo*, we injected mice with each blocking antibody over eight days (Fig. 2a). Anti-JAG1.b70 but not anti-JAG2.b33 induced a significant decrease in club cells and a corresponding increase in ciliated cells, as assessed by immunofluorescent detection of club and ciliated cell markers (Fig. 2b, c and Extended Data Fig. 4a). The lack of an altered lung phenotype after JAG2 inhibition was not due to a lack of antibody activity, because anti-JAG2.b33, but not anti-JAG1.b70, caused a near-complete loss of sebaceous gland cells, providing pharmacodynamic evidence of JAG2 antagonism (Extended Data Fig. 4b). This dominance of JAG1 versus JAG2 fits with cell fate analyses during lung development[4]. Increasing the dosage (to $30\,\mathrm{mg\,kg^{-1}}$) or treatment duration (to several weeks) did not alter the phenotypes, bolstering the conclusion that each antibody is selective *in vivo* for its cognate ligand, and indicating that systemic JAG1 or JAG2 blockade is well tolerated (data not shown).

Notably, dual blockade of JAG1 plus JAG2 generated bronchiolar epithelium that was nearly devoid of club cells and instead almost completely comprised of ciliated cells, as assessed by immunofluorescence, quantification and electron microscopy (Fig. 2b–d and Extended Data Fig. 4a). Intranasal antibody delivery also induced this phenotype, consistent with a direct effect on the epithelium (Extended Data Fig. 4c). Although the loss of club cells appeared nearly complete, rare cells remained positive for the club cell marker CC10 (also known as SCGB1A1) at two locations: adjacent to pulmonary neuroendocrine cells (Extended Data Fig. 4d) and at bronchoalveolar duct junctions (Extended Data Fig. 8a). We speculate that these may be variant club cells proposed to participate in epithelial regeneration[18,19]. Transcriptomic analysis confirmed that our JAG antagonists inhibited Notch signalling, and revealed changes in lineage marker expression consistent with histological analyses

[1]Department of Discovery Oncology, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA. [2]Department of Antibody Engineering, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA. [3]Department of Structural Biology, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA. [4]Department of Pathology, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA. [5]Department of Translational Immunology, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA. [6]Department of Discovery Immunology, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA. [7]Department of Molecular Biology, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA. [8]Departments of Protein Chemistry, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA. †Present addresses: Department of Biomedical Engineering, University of California, Davis, 451 E. Health Sciences Drive, Davis, California 95616, USA (C.S.); Amgen, Inc., 1120 Veterans Boulevard, South San Francisco, California 94080, USA (J.E.).

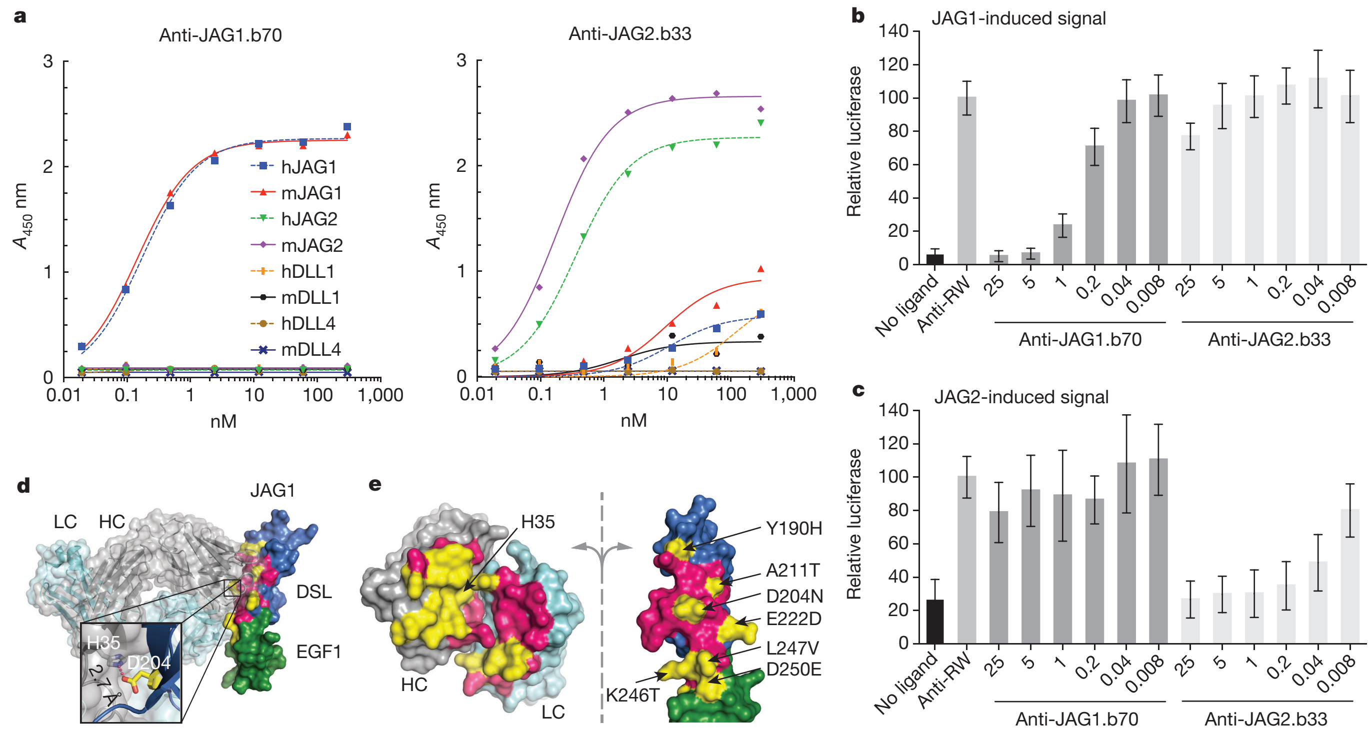


Figure 1 | Anti-JAG1.b70 and anti-JAG2.b33 specifically block signalling induced by their cognate ligands. a, ELISA measuring antibody binding to purified protein fragments of human (h) and murine (m) canonical Notch ligands. **b**, **c**, Reporter assays of Notch signalling induced by JAG1 (**b**) or JAG2 (**c**); antibody concentrations are shown in $\mu g\,ml^{-1}$ ($25\,\mu g\,ml^{-1}$ corresponds to 170 nM); values reflect reporter gene signalling relative to control reporter (mean ± s.d., $n = 12$ from three independent experiments with four replicates in each experiment); anti-ragweed (RW), isotype control antibody at $25\,\mu g\,ml^{-1}$. **d**, Side-view of anti-JAG1 Fab bound to the human JAG1 DSL–EGF1 region. JAG1 residues within 4.5 Å of the Fab fragment (HC, heavy chain; LC, light chain) paratope are pink or yellow, with yellow designating residues unique to JAG1, relative to JAG2. Inset shows an ionic interaction between Asp204 (hJAG1) and His35 (HC) that anchors the JAG1–Fab interaction. **e**, Open-book view of the complex. JAG1 is coloured as in **d**, with highlighted residues named according to side-chain identity in human JAG1 and JAG2, respectively.

(Extended Data Fig. 4e). Antagonist antibodies targeting NOTCH1 and NOTCH2 (ref. 20) revealed NOTCH2 as the dominant receptor, with a secondary role for NOTCH1 (Extended Data Fig. 5a), consistent with studies during development[5]. Adding to our analyses of ligand and receptor expression (Extended Data Fig. 5b–e), these results demonstrate that active Notch signalling is required for club cell maintenance, with JAG1 on ciliated cells and NOTCH2 on club cells as the dominant pairing.

Notably, we detected obvious club cell loss as early as four days after antibody dosing, with the phenotypic switch complete by day 6 (Fig. 2e), significantly accelerated relative to homeostatic turnover of 6–12 months[11]. We considered that JAG blockade might induce cell death, followed by progenitor proliferation and differentiation. However, we detected no increase in apoptosis or proliferation in bronchiolar epithelial cells, as assessed by staining for cleaved caspase (data not shown) or KI67 (Fig. 2f and Extended Data Fig. 6a). We also included BrdU throughout antibody dosing to label any cell that proliferated during epithelial conversion (Fig. 2g). While we observed a slight increase in the percentage of labelled cells after JAG blockade, we did not detect proliferation in club cells (see below), and this percentage (2.2%) is insufficient to consider proliferation as a driving force behind the near-complete cell fate switch (Fig. 2h and Extended Data Fig. 6b).

This rapid-onset phenotype in the absence of notable proliferation raised the provocative possibility that club cells were directly converting into ciliated cells. Consistent with this hypothesis, JAG blockade induced the appearance of 'intermediate' cells that expressed markers of both lineages (Extended Data Fig. 6c–f). To trace changes in club cell fate, we generated a Scgb1a1-CreERT2^{GNE} mouse line (Extended Data Fig. 7a) and crossed it to the Rosa26-lsl-tdTomato reporter strain[21]. Induced club cell labelling (Fig. 3a) was detected in 2.4% of airway epithelial cells, and this frequency was not altered by JAG blockade (Extended Data Fig. 7b), supporting the conclusion that JAG inhibition does not prompt club cell death or proliferation. In control epithelium, the vast majority of tdTomato-positive cells expressed CC10 (97.80% ± 2.20% (mean ± s.d.)), whereas only a small percentage expressed the ciliated cell marker FOXJ1 (5.00% ± 1.70%) (Fig. 3b, c), consistent with rare intermediate cells in control lung. In obvious contrast, six days of JAG or NOTCH inhibition (Fig. 3b, c and Extended Data Fig. 7c, d) reversed the fate distribution, such that the vast majority of tdTomato-positive cells now expressed ciliated cell markers acetylated-α-tubulin or FOXJ1 (85.5% ± 2.8%, Fig. 3b, c), whereas only 11.7% ± 2.9% of them expressed CC10 (Fig. 3c). The newly generated ciliated cells were functional, as determined by examining cilia motility from live tdTomato-positive cells isolated after JAG blockade (Supplementary Videos 1 and 2). JAG blockade did not increase BrdU incorporation in traced cells above the low level (<0.02%) found in controls (Fig. 3d), demonstrating that JAG inhibition does not increase division of the relevant cell population. Our lineage-tracing experiments thus establish that JAG inhibition induces a direct transdifferentiation of one cell type to another, without cell division.

Thirteen weeks after a single antibody dose, the distal airways had largely recovered, although the more proximal airways remained ciliated (Extended Data Fig. 8a). This slow rate cannot be explained by antibody perdurance, because measurements of antibody half-life (approximately one week) informed us that antibody serum levels fell below the efficacious threshold after three weeks (data not shown). After a 13-week chase under our CC10 lineage-tracing conditions (Fig. 3a), and as expected for a slowly regenerating organ, a significant fraction of traced cells remained as single cells, both after control (79.50% ± 3.25%) and anti-JAG (84.40% ± 5.28%) treatments (Fig. 3e). In controls, the vast majority of these single cells were club cells (92.19% ± 4.48%), whereas 91.54% ± 7.10% of such cells remained ciliated after JAG blockade (Fig. 3f). In the control group, expanded

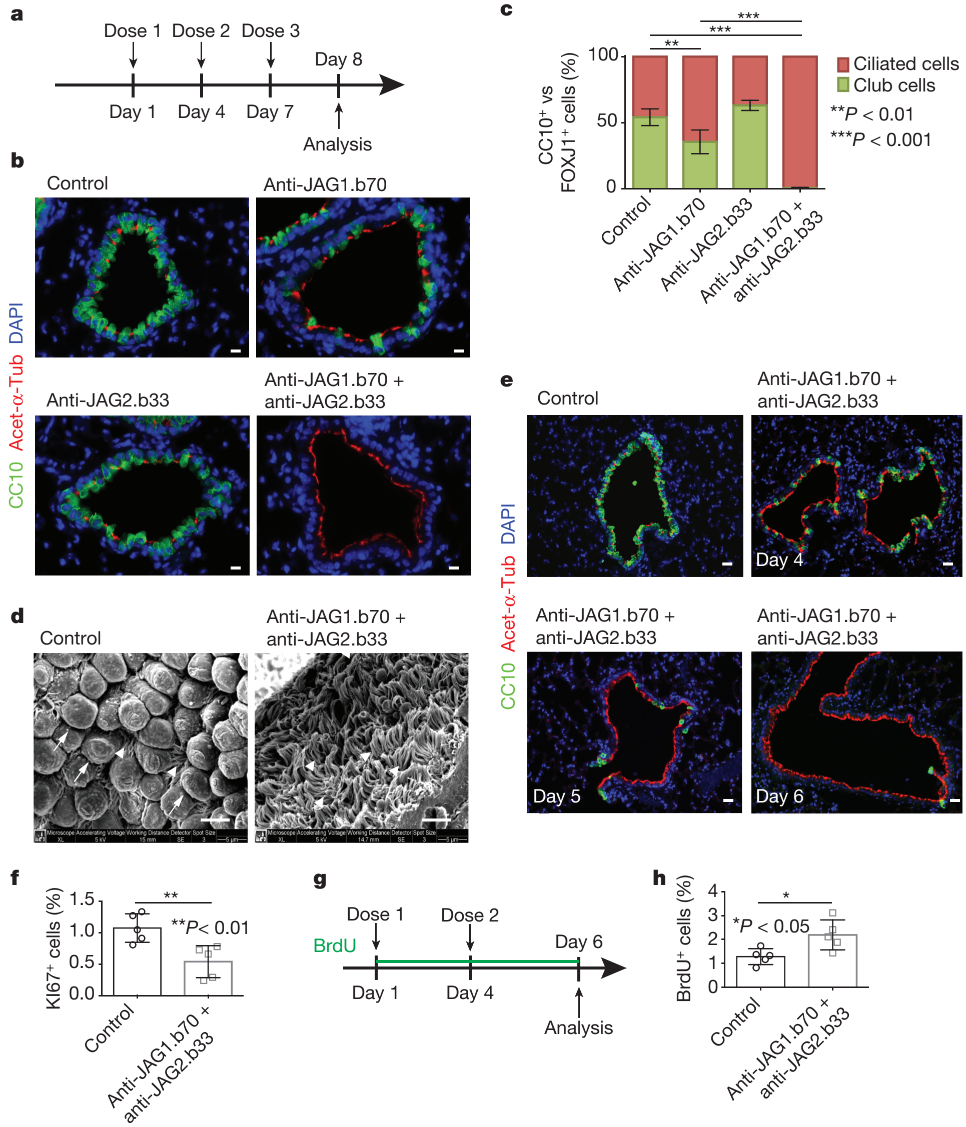


Figure 2 | JAG blockade induces rapid near-complete club cell loss and ciliated cell gain without proliferation. a, Schematic of antibody treatment schedule. **b**, Immunofluorescent staining for the club cell marker CC10 (green) and the ciliated cell marker acetylated-α-tubulin (acet-α-Tub; red) in distal airway epithelium. Nuclei counterstained with DAPI. **c**, Quantification of club and ciliated cell numbers as assessed by immunofluorescent staining for CC10 and FOXJ1, respectively (percentages of each cell type over the total number of airway cells counted, $n = 5$; unpaired t-test, $^{**}P < 0.01$, $^{***}P < 0.001$). **d**, Scanning electron microscopy, revealing an epithelium rich in club cells (arrows) after control treatment (left), in contrast to an epithelium made exclusively of ciliated cells (arrowheads) after JAG blockade (right). **e**, Kinetics of cell fate changes. Mice were treated with a single antibody dose and analysed on the indicated days by immunofluorescence for club cells or ciliated cells as in **b** ($n = 3$ mice per time point). **f**, Quantification of the percentage of airway cells stained for the proliferation marker KI67 (white) as assessed by immunofluorescence ($n = 5$ mice per group; unpaired t-test, $^{**}P < 0.01$). **g**, Mice were dosed twice with antibodies over six days with 1 mg ml^{-1} BrdU in the drinking water to assess cell division. **h**, Quantification of the percentage of airway cells positive for BrdU immunofluorescence staining ($n = 5$ mice per group; unpaired t-test, $^{*}P < 0.05$). Data are mean ± s.d. Scale bars, 10 μm (**b**), 5 μm (**d**) and 20 μm (**e**).

clones were small (two to four cells), but after JAG blockade the clones were larger (average of seven cells) (Fig. 3e). These clones localized at bronchoalveolar duct junctions, where rare club cells remained after JAG blockade, and contained both club and ciliated cells (Extended Data Fig. 8b–d). We propose that the transdifferentiated ciliated cells do not 'reconvert' to club cells but instead are replaced during normal epithelial turnover with directionality, from the bronchoalveolar duct junctions to the larger airways.

Excess mucus secretion in airways is a unifying complication of several diseases, including asthma, idiopathic pulmonary disease and chronic obstructive pulmonary disease. Thus, we tested whether NOTCH blockade could convert mucus-secreting to mucus-clearing cells in an oft-studied pre-clinical model of goblet cell metaplasia[22]. Sensitized mice were challenged with inhaled ovalbumin to induce inflammation, which stimulates club cells to differentiate into goblet cells[23,24], the source of aberrant mucus secretion. In a prevention study, with antibody delivered 24 h after the first ovalbumin inhalation, we found that blocking a JAG1–NOTCH2 signalling axis reduced goblet cell metaplasia to that of non-sensitized animals (Extended Data Fig. 9, controls in Fig. 4b, c). Notably, we found that anti-JAG1.b70—either alone or in combination with anti-JAG2.b33—also effectively reversed goblet cell metaplasia that had been fully established before antibody dosing (Fig. 4). Neither ligand nor receptor inhibition effected the inflammation severity score, our standard method for assessing the immune response (Fig. 4h and Extended Data Fig. 9d, g). Likewise, JAG blockade did not significantly reduce the number of lung eosinophils, the main drivers of lung inflammation in this model[25,26] (Extended Data Fig. 10a, b), nor did it alter cytokines in a manner that could explain a possible anti-inflammatory effect of antibody treatment (Extended Data Fig. 10c, d). These results point to

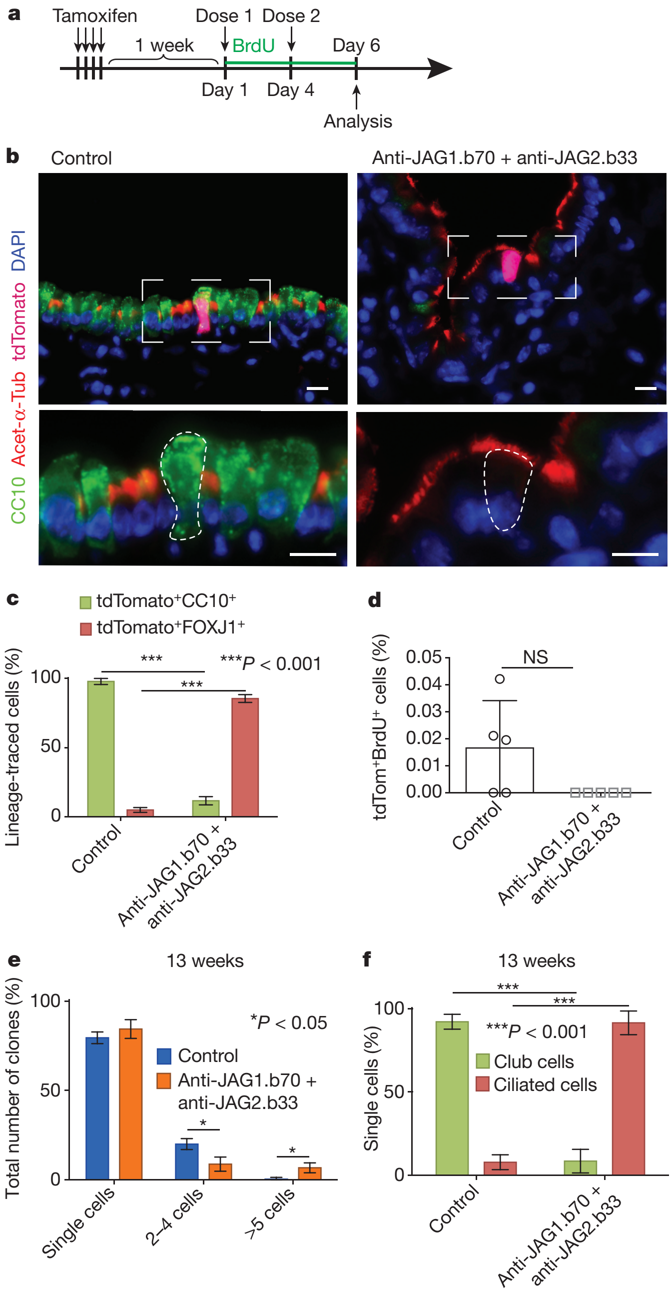


Figure 3 | Lineage tracing demonstrates that club cells transdifferentiate into ciliated cells, which do not revert back after antibody washout. **a**, Schematic of Tamoxifen induction and antibody treatment of Scgb1a1-CreERT2^{GNE}/Rosa26-lsl-tdTomato mice. **b**, Immunofluorescence staining of lineage-traced cells for CC10 (green), acetylated-α-tubulin (red) and the lineage-tracing marker tdTomato (pink). Boxed areas are enlarged in the bottom row, with tdTomato-positive cells outlined but not coloured ($n = 5$ mice per group). Nuclei counterstained with DAPI. Scale bars, 10 μm. **c**, Quantification of the percentage of lineage-traced cells (tdTomato$^+$) expressing club (CC10) or ciliated (FOXJ1) cell markers ($n = 5$ mice per group, mean ± s.d.; unpaired t-test, $^{***}P < 0.001$). **d**, Quantification of the percentage of proliferating lineage-traced cells (tdTomato$^+$/BrdU$^+$) in the lung epithelium ($n = 5$ mice per group, mean ± s.d.; unpaired t-test, NS, not significant). **e**, Analysis of clonal expansion of lineage traced club cells during the recovery period, 13 weeks after a single antibody dose. **f**, Cell fate analysis of single traced club cells from **e**, based on immunofluorescence staining for CC10 (club cells) and FOXJ1 (ciliated cells).

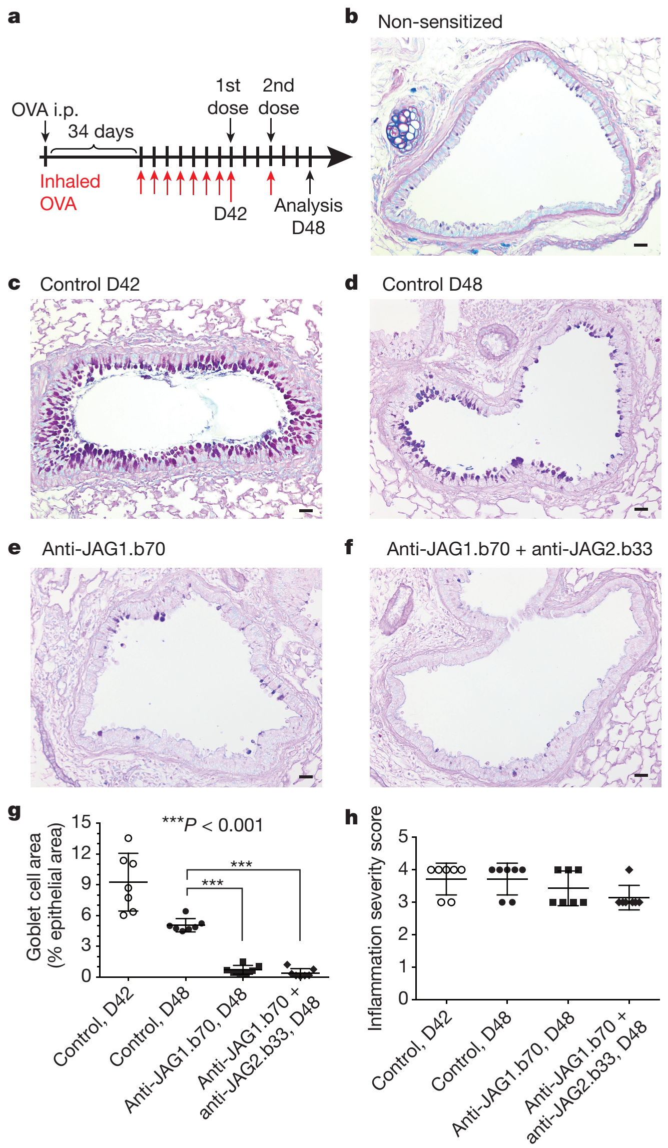


Figure 4 | JAG blockade reverses goblet cell metaplasia *in vivo*. **a**, Mice were sensitized during a 34-day period after intraperitoneal (i.p.) injection of ovalbumin (OVA) or vehicle (non-sensitized, **b**) and then challenged with inhaled ovalbumin to induce inflammation and goblet cell metaplasia. D, day. **c**–**f**, Alcian blue/periodic acid-Schiff (PAS) staining for mucin from lung sections of sensitized mice treated with control antibody at the first (day 42) (**c**) or second (day 48) (**d**) dose of control antibody, or at the second dose of anti-JAG1.b70 (**e**) or anti-JAG1.b70 plus anti-JAG2.b33 (**f**). Scale bars, 20 μm. **g**, Quantification of goblet cell area ($n = 7$ mice per group, mean ± s.d., unpaired t-test, $^{***}P < 0.001$). **h**, Inflammation index as assessed by haematoxylin and eosin staining.

an epithelial-cell-specific mechanism, and suggest that prevention and reversal of goblet cell metaplasia reflect direct effects of JAG inhibition on lung cell fate.

By acutely blocking the Notch pathway, we have uncovered an inherent requirement for Notch activity in club cells to maintain their cell fate in the adult homeostatic lung. Pharmacological inhibition of Jagged ligands delineates a previously uncharacterized relationship between club and ciliated cells, with Notch signalling induced in the former by Jagged expression on the latter. Disrupting this interaction unveiled a remarkable plasticity of adult club cells, which transdifferentiated into ciliated cells within a few days without cell division. At the same time, previous studies of the club-ciliated cell relationship under proliferative conditions perhaps portended our findings[27–29]. Club cells have been referred to as a 'not undifferentiated' cell type, acknowledging the paradoxical combination of specialized function that is a hallmark of a terminally differentiated cell together with an ability to generate other lineages through division and differentiation[29]. Recent genetic

studies suggest that other cell types may also regulate this cell fate in the trachea, which includes a basal layer proposed to be the key source of Jagged required for Notch activation in the differentiated club cells and proliferating progenitors[8,30].

Our antibodies enable pharmacological conversion of a diseased airway from an epithelium aberrantly producing mucus to one primarily but temporarily functioning to clear mucus, with hope for a therapeutic window in the absence of toxicities associated with pan-NOTCH inhibition. This ability to quickly modulate cell fate thus holds promise for a new type of therapy.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 February; accepted 10 September 2015.
Published online 18 November 2015.

1. Koch, U., Lehal, R. & Radtke, F. Stem cells living with a Notch. *Development* **140,** 689–704 (2013).
2. Chilvers, M. A. & O'Callaghan, C. Local mucociliary defence mechanisms. *Paediatr. Respir. Rev.* **1,** 27–34 (2000).
3. Roy, M. G. *et al.* Muc5b is required for airway defence. *Nature* **505,** 412–416 (2014).
4. Zhang, S., Loch, A. J., Radtke, F., Egan, S. E. & Xu, K. Jagged1 is the major regulator of Notch-dependent cell fate in proximal airways. *Dev. Dyn.* **242,** 678–686 (2013).
5. Morimoto, M., Nishinakamura, R., Saga, Y. & Kopan, R. Different assemblies of Notch receptors coordinate the distribution of the major bronchial Clara, ciliated and neuroendocrine cells. *Development* **139,** 4365–4373 (2012).
6. Rock, J. R. *et al.* Notch-dependent differentiation of adult airway basal stem cells. *Cell Stem Cell* **8,** 639–648 (2011).
7. Pardo-Saganta, A. *et al.* Injury induces direct lineage segregation of functionally distinct airway basal stem/progenitor cell subpopulations. *Cell Stem Cell* **16,** 184–197 (2015).
8. Mori, M. *et al.* Notch3-Jagged signaling controls the pool of undifferentiated airway progenitors. *Development* **142,** 258–267 (2015).
9. Kotton, D. N. & Morrisey, E. E. Lung regeneration: mechanisms, applications and emerging stem cell populations. *Nature Med.* **20,** 822–832 (2014).
10. Hogan, B. L. M. L. M. M. L. M. *et al.* Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell* **15,** 123–138 (2014).
11. Rawlins, E. L. & Hogan, B. L. M. Ciliated epithelial cell lifespan in the mouse trachea and lung. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **295,** L231–L234 (2008).
12. Rock, J. R. & Hogan, B. L. M. Epithelial progenitor cells in lung development, maintenance, repair, and disease. *Annu. Rev. Cell Dev. Biol.* **27,** 493–512 (2011).
13. Watson, J. K. *et al.* Clonal dynamics reveal two distinct populations of basal cells in slow-turnover airway epithelium. *Cell Rep.* **12,** 90–101 (2015).
14. Rawlins, E. L. & Hogan, B. L. M. Epithelial stem cells of the lung: privileged few or opportunities for many? *Development* **133,** 2455–2465 (2006).
15. Martin, C., Frija-Masson, J. & Burgel, P. R. Targeting mucus hypersecretion: New therapeutic opportunities for COPD? *Drugs* **74,** 1073–1089 (2014).
16. Luca, V. C. *et al.* Structural basis for Notch1 engagement of Delta-like 4. *Science* **347,** 847–853 (2015).
17. Cordle, J. *et al.* A conserved face of the Jagged/Serrate DSL domain is involved in Notch *trans*-activation and *cis*-inhibition. *Nature Struct. Mol. Biol*. **15,** 849–857 (2008).
18. Kim, C. F. B. *et al.* Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* **121,** 823–835 (2005).
19. Guha, A. *et al.* Neuroepithelial body microenvironment is a niche for a distinct subset of Clara-like precursors in the developing airways. *Proc. Natl Acad. Sci. USA* **109,** 12592–12597 (2012).
20. Wu, Y. *et al.* Therapeutic antibody targeting of individual Notch receptors. *Nature* **464,** 1052–1057 (2010).
21. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neurosci.* **13,** 133–140 (2010).
22. Misaghi, S. *et al.* Polyclonal hyper-IgE mouse model reveals mechanistic insights into antibody class switch recombination. *Proc. Natl Acad. Sci. USA* **110,** 15770–15775 (2013).
23. Chen, G. *et al.* SPDEF is required for mouse pulmonary goblet cell differentiation and regulates a network of genes associated with mucus production. *J. Clin. Invest.* **119,** 2914–2924 (2009).
24. Pardo-Saganta, A., Law, B. M., Gonzalez-Celeiro, M., Vinarsky, V. & Rajagopal, J. Ciliated cells of pseudostratified airway epithelium do not become mucous cells after ovalbumin challenge. *Am. J. Respir. Cell Mol. Biol.* **48,** 364–373 (2013).
25. Shin, Y. S., Takeda, K. & Gelfand, E. W. Understanding asthma using animal models. *Allergy Asthma Immunol. Res.* **1,** 10–18 (2009).
26. Nials, A. T. & Uddin, S. Mouse models of allergic asthma: acute and chronic allergen challenge. *Dis. Model. Mech*. **1,** 213–220 (2008).
27. Xing, Y., Li, A., Borok, Z., Li, C. & Minoo, P. NOTCH1 is required for regeneration of Clara cells during repair of airway injury. *Stem Cells* **30,** 946–955 (2012).
28. Zheng, D., Yin, L. & Chen, J. Evidence for Scgb1a1$^+$ cells in the generation of p63$^+$ cells in the damaged lung parenchyma. *Am. J. Respir. Cell Mol. Biol.* **50,** 595–604 (2014).
29. Rawlins, E. L. *et al.* The role of Scgb1a1$^+$ Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell Stem Cell* **4,** 525–534 (2009).
30. Pardo-Saganta, A. *et al.* Parent stem cells can serve as niches for their daughter cells. *Nature* **523,** 597–601 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank E. Jackson and R. Pattni for work generating the Scgb1a1-ERT2^{GNE} mice; L.Nguyen, L. Orellana, P. Grigg and the Genentech Transgenic Technology Laboratories and Research Support Facility for technical assistance with mouse strains and colonies; F. Chu, L. Rangell, S. Chalasani, C. Jones III and C. Espiritu for cell staining; A. Ertürk, C. Chalouni, S. Gierke, M. Gonzalez-Edick and the Genentech Center for Advanced Light Microscopy (CALM) for imaging; C. K. Poon for cytokine measurements; S. P. Tsai and M. Dostalek for pharmacokinetic analyses; T. Hagenbeek for help with the immune cell studies. Use of the Stanford Synchrotron Radiation Lightsource SSRL 12-12 at Stanford Linear Accelerator Center National Accelerator Laboratory is supported by the US Department of Energy (DOE), DOE Office of Biological and Environmental Research, National Institutes of Health, and National Institute of General Medical Sciences. The contents of this publication are the responsibility of the authors and do not necessarily represent the views of NIH or NIGMS.

Author Contributions D.L. performed experiments and analysed data of Figs 2–4 and Extended Data Fig. 4–10, A.S. performed reporter assays for Fig. 1 and experiments for Extended Data Fig. 2c and mouse studies for Extended Data Fig. 4b, G.d.L.B. purified antibody fragments and crystallized the JAG1–Fab complex, Y. Chen and S.S.S performed affinity maturation and characterization of antibodies, C.C., S.W. and Y.W. generated the phage display antibodies and performed the *in vitro* binding experiments and affinity maturation, M.R. performed all electron microscopy studies, C.S. performed the cilia functionality studies, M.R.-G. and S.W. designed the targeting vector and supervised the generation of the Scgb1a1-CreERT2^{GNE} mouse line, M.Z. and X.Wu performed the ovalbumin studies, J.E.-A. performed all quantifications of immunofluorescence staining, H.M. performed qPCR and analysis of whole lungs, Y.Chinn and J.Q.H. assisted with anti-JAG1 development and expressed and purified JAG1 protein; W.P.L. helped design and supervise the ovalbumin study, C.A. analysed tissue sections from the ovalbumin study, J.E. contributed to the design of qPCR and ovalbumin studies as well as contributing intellectually, J.P. solved and analysed the structure, and made the structural figures in Fig. 1 and Extended Data Fig. 3, J.B.L. analysed the histology of all lung and skin samples except samples from the ovalbumin study. C.W.S. supervised the experiments and wrote the paper with D.L.

Author Information Atomic coordinates and structure factors for the human JAG1–Fab complex have been deposited in the Protein Data Bank under accession code 5BO1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.W.S. (csiebel@gene.com).

METHODS

Phage library sorting and screening. Recombinant human JAG1 (extracellular domain (ECD)) and human and mouse JAG2 (DSL–EGF1–4) were used as antigens for library sorting. Nunc 96-well Maxisorp immunoplates were coated overnight at 4 °C with target antigen (10 μg ml^{-1}) and blocked for 1 h at room temperature with PBST buffer (PBS with 1% (w/v) bovine serum albumin (BSA) and 0.05% (v/v) Tween-20). Either VH antibody phage libraries[31] or VH/VL libraries[32] were added to antigen plates and incubated overnight at room temperature. Plates were washed ten times with PBT (PBS with 0.05% Tween-20), and bound phage were eluted with 50 mM HCl plus 500 mM NaCl for 30 min, and neutralized with an equal volume of 1 M Tris base, pH 7.5. Recovered phages were amplified in *Escherichia coli* XL-1 Blue cells. During subsequent selection rounds, phage-antigen incubation times were reduced to 2–3 h, while the stringency of plate washing was gradually increased. After four rounds, 96 clones were picked from each of the VH and VH/VL library sorts and analysed for JAG1 and JAG2 binding. The phage supernatant was diluted 1:5 in ELISA buffer (PBS with 0.5% BSA, 0.05% Tween-20) in 100 μl total volume and transferred to plates coated with target protein (1 μg ml^{-1} directly coated overnight). After 1 h of gentle shaking to allow phage binding, the plate (room temperature) was washed ten times with PBST, incubated for 30 min with horseradish peroxidase (HRP)-conjugated anti-M13 antibody in ELISA buffer (1:5,000), washed ten times with PBST, and incubated for 5 min with 50 μl each of 3,3′,5,5′-tetramethylbenzidine (TMB) peroxidase substrate and peroxidase solution B (H_2O_2) (KirkegaardPerry Laboratories). Reactions were stopped with 100 μl 0.1 M phosphoric acid (H_3PO_4) and absorbance was determined at 450 nm. The reduction in absorbance (%) was calculated by the following equation: $A_{450\ nm}$ reduction (%) = [($A_{450\ nm}$ of wells with competitor)/($A_{450\ nm}$ of wells without competitor)] × 100. Clones that bound specifically, defined as yielding an $A_{450\ nm}$ at least fivefold higher for JAG1 or JAG2 binding over background, were selected and JAG1–JAG2 cross-binders were eliminated. Unique V_L and V_H sequences were cloned into the LPG3 and LPG4 vectors, respectively, to generate full-length human IgG1 constructs, which were expressed in mammalian CHO cells and purified with a protein A column.

Affinity improvement. For each clone that showed promising cell-based assay activity, phagemid containing four stop codons (TAA) in complementarity determining region loop 3 (CDR L3) and displaying monovalent Fab on the surface of M13 bacteriophage was generated. These phagemids served as the templates for Kunkel mutagenesis for the construction of affinity maturation libraries. For affinity maturation, soft randomization strategy was used, where mutagenic DNA was synthesized with 70–10–10–10 mixtures of bases favouring the wild-type nucleotides to obtain the mutation rate of approximately 50% at the selected positions. Four different combinations of CDR loops, H1/L3, H2/L3, H3/L3 and L1/L2/L3 were selected for randomization.

For affinity improvement selection, human JAG1 or JAG2 was first biotinylated under limiting reagent condition. Phage libraries were subjected to six rounds of solution sorting with increasing stringency. For the first round of solution sorting, three absorbance units per ml in 1% BSA and 0.05% Tween-20 of phage input were incubated to plates pre-coated with antigens for 3 h. The wells were washed with PBT ten times. Bound phage was eluted with 150 μl per well of 50 mM HCl and 500 mM KCl for 30 min, and subsequently neutralized by 50 μl per well of 1 M Tris, pH 8.0, titred, and propagated for the next round. For subsequent rounds, panning of the phage libraries was done in solution phase, in which phage library was incubated with initial concentration of 200 nM for biotinylated JAG1 and 20 nM for biotinylated JAG2 protein (the concentration is based on parental clone equilibrium dissociation constant) in 100 μl buffer containing 1% Superblock (Pierce Biotechnology) and 0.05% Tween-20 for 2 h at room temperature. The mixture was further diluted ten times with 1% Superblock, and 100 μl per well was applied to neutravidin-coated wells (10 μg ml^{-1}) for 30 min at room temperature with gentle shaking. To determine the background binding, control wells containing phage were captured on neutravidin-coated plates. Bound phage was then washed, eluted and propagated as described for first round. Five more rounds of solution sorting were carried out together with increasing selection stringency. The first couple rounds of which is for on-rate selection by decreasing biotinylated target protein concentration from 200 nM or 20 nM to 0.5 nM, and the last two rounds of which is for off-rate selection by adding excess amounts of non-biotinylated target protein (300–1,000-fold more) to compete off weaker binders at room temperature.

Affinity screening ELISA (single spot competition). Colonies were picked from the sixth round of screening and were grown overnight at 37 °C in 1.50 ml per well of 2YT media with 50 μg ml^{-1} carbenicillin and 1×10^{10} per ml M13KO7 in 96-well plate (Falcon). From the same plate, a colony of XL-1-infected parental phage was picked as control. 96-well Nunc Maxisorp plates were coated with 100 μl per well of either JAG1 or JAG2 (0.5 μg ml^{-1}) in PBS at 4 °C overnight. The plates were blocked with 150 μl of 1% BSA in PBST for 1 h.

The phage supernatant (35 μl) was diluted with 75 μl ELISA buffer with or without 25 nM JAG1 or 5 nM JAG2, and incubated for 1 h at room temperature in an F plate (NUNC). The mixture (95 μl) was transferred side by side to the antigen-coated plates. The plate was gently shaken for 15 min and washed ten times with PBT. The binding was quantified by adding HRP-conjugated anti-M13 antibody in ELISA buffer (1:2,500) and incubated for 30 min at room temperature. The plates were washed with PBST ten times. Next, 100 μl per well of peroxidase substrate was added to the well and incubated for 5 min at room temperature. The reaction was stopped by adding 100 μl 0.1 M phosphoric acid (H_3PO_4) to each well and allowed to incubate for 5 min at room temperature. The absorbance of the yellow colour in each well was determined using a standard ELISA plate reader at 450 nm. In comparison to the $A_{450\ nm}$ reduction (%) of the well of parental phage (100%), clones that had the $A_{450\ nm}$ reduction (%) lower than 50% were picked for sequence analysis. Unique clones were selected for phage preparation to determine binding affinity (phage IC_{50}) against target antigen by comparison to parental clone. Clone that showed most affinity improvement were reformatted into human IgG1 for antibody production and further BIAcore binding kinetic analysis and other *in vitro* or *in vivo* assays.

Antibody binding assays. JAG1 and JAG2 antibodies were tested for binding to recombinant purified Notch ligands human JAG1 (hJAG1), mouse JAG1 (mJAG1), human JAG2 (hJAG2), murine JAG2 (mJAG2), human DLL1 (hDLL1), mouse DLL1 (mDLL1), human DLL4 (hDLL4) and mouse DLL4 (mDLL4) using a standard ELISA. Notch ligand protein (1 μg ml^{-1}) in PBS, pH 7.4, was coated on ELISA plates (Nunc Maxisorp) at 4 °C overnight. Plates were blocked with casein blocker in PBS (Pierce) for 1 h at room temperature. Serial threefold dilutions of antibody IgGs in PBST buffer were added to the plates and incubated for 1 h at room temperature. The plates were then washed with PBST and bound antibodies were detected with peroxidase-conjugated goat anti-human Fab specific IgG (Sigma). TMB substrate (3,3′,5,5′-tetramethylbenzidine) was used and the reactions were stopped with 100 μl 0.1 M phosphoric acid (H_3PO_4) before absorbance at 450 nM was read using a standard ELISA plate reader. Absorbance was plotted against concentrations of IgGs using Prism 6 (GraphPad Software).

Antibody binding affinities. Antibody binding affinities and rate constants were measured by surface plasmon resonance using a BIAcore-T200 instrument. For human JAG1, human and mouse JAG2, human DLL1, human and mouse DLL4 affinity measurements, human IgG versions of anti-JAG1.b70 and anti-JAG2.b33 antibodies were captured by mouse anti-human Fc antibody (GE Healthcare, BR-1008-39) coated on CM5 biosensor chips to achieve approximately 200 response units (RU). For kinetics measurements, fourfold serial dilutions (480–0.117 nM) of ligands were injected in HBS-T buffer (0.01 M HEPES, pH 7.4, 0.15 M NaCl, 0.05% (v/v) Surfactant P20, GE Healthcare) at 25 °C with a flow rate of 30 μl min^{-1}. For mouse JAG1 and DLL1, owing to their background binding to empty flow cell, the kinetic parameters were determined via directly coating the ligands on the CM5 biosensor chip. Purified antibodies in fragment antigen-binding (Fab fragment) format were then flown through the biosensor chip. Mouse JAG1 or DLL1 was coated on CM5 biosensor chips to achieve approximately 100 RU. Fivefold serial dilutions (500–0.16 nM) of Fab fragments were then injected in HBS-T at 25 °C with a flow rate of 30 μl min^{-1}. Association (k_{on}) and dissociation (k_{off}) rates were calculated using a simple one-to-one Langmuir binding model (BIAcore Evaluation T200 Software version 2.0). The equilibrium dissociation constant (K_D) was calculated as the ratio k_{off}/k_{on}. For affinity analysis, K_d was calculated using a steady state affinity model.

Notch reporter assays. U87 glioblastoma cells, which endogenously express predominantly NOTCH2 but only very low levels of other NOTCH receptors, were co-transfected with a Notch-responsive TP-1 (12X CSL) firefly luciferase reporter and a constitutively expressed *Renilla* luciferase reporter (pRL-CMV, Promega E2261) to control for transfection efficiency. Antibodies were added with the ligand-expressing cells (NIH-3T3 cells stably transfected with human JAG1 or JAG2 6–8 h after transfection). Luciferase activities were measured after 20 h of co-culture (Dual Glo Luciferase, Promega E2920), using a Perkin-Elmer EnvVison 2103 Multilabel Reader. Typically, four replicates were analysed for each condition, and values were expressed as relative luciferase units (firefly signal divided by the *Renilla* signal) and graphed as percentage of signalling relative to anti-ragweed isotype control antibody, which was set at 100%. The cell line tested negative for mycoplasma.

NOTCH2 intracellular domain immunoblot analysis. NOTCH2 signalling was induced in U87 glioblastoma cells by incubation with JAG1-coated (R&D, 599-JG) beads (Bangs Laboratories, BM562) for 24 h. Cells were subsequently collected and nuclear fractions were isolated, and 10 μg of protein was run on a 4–12% NuPAGE Novex Bis-Tris gel buffered with MOPS (Life technologies) for 90 min at a constant voltage of 200 V. Proteins were then transferred to PVDF membrane. The NOTCH2 intracellular domain was detected using our in-house antibody

(clone 40-2-7)[33] at a concentration of 0.2 μg ml^{-1}. The nuclear protein CREB (Cell Signaling 9197, clone 48H2) was detected as a loading control.

JAG1, JAG2, DLL1 and DLL4 fragment reagent proteins. Mouse DLL1 (M1-Q516-CHis) and human DLL1 (M1-G540-CHis) were purchased from Sino Biological (50522-M08H and 11635-H08H-50, respectively). Mouse DLL4 (S28-P525-CHis) and human DLL4 (S27-P524-CHis) were purchased from R&D systems (1389-D4 and 1506-D4, respectively). Mouse JAG1 (M1-D387-CHis), human JAG1 (M1-D387-CHis), mouse JAG2 (M1-E403-CHis) and human JAG2 (M1-E403-CHis) were cloned into a modified pAcGP67A vector and all constructs were confirmed by DNA sequencing. Recombinant baculovirus was generated using the Baculogold system (BD Biosciences) following standard protocols. The virus was amplified twice to prepare the stock used for protein expression. Protein was expressed in Sf9 cells in serum-free ESF921 (Expression Systems LLC) and cells were grown to 2E6 cells per ml and infected with the appropriate virus at a ratio of virus/culture of 5.0 ml l^{-1} (v/v). The Wave reactor was maintained at 27 °C and 25 r.p.m. with fixed angle of 9° and 0.3 l m^{-1} of 30% oxygen. At 48 h after infection, cells were pelleted by centrifugation at 5,000*g* for 15 min. The culture media were supplemented with 50 mM Tris, pH 8, 5 mM $CaCl_2$ and 1 mM $NiCl_2$ and proteins were purified over a Ni-NTA Superflow column (Qiagen), washed with 10 column volumes of buffer P1 (20 mM Tris, pH 8, 300 mM NaCl and 30 mM imidazole), eluted with buffer P2 (20 mM Tris, pH 8, 300 mM NaCl and 300 mM imidazole), and further purified over a Superdex 75 column (GE Healthcare) in buffer P3 (50 mM HEPES, pH 7.5, 200 mM NaCl and 5 mM $CaCl_2$). Peak fractions were analysed by SDS–PAGE, pooled, aliquoted and frozen at −80 °C. Protein sequences of these constructs with signal sequences and purification tags underlined were mouse JAG1: MLLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFAADLGSQFELEILSMQNVNGELQNGNCCGGVRNPGDRKCTRDECDTYFKVCLKEYQSRVTAGGPCSFGSGSTPVIGGNTFNLKASRGNDRNRIVLPFSFAWPRSYTLLVEAWDSSNDTIQPDSIIEKASHSGMINPSRQWQTLKQNTGIAHFEYQIRVTCDDHYYGFGCNKFCRPRDDFFGHYACDQNGNKTCMEGWMGPDCNKAICRQGCSPKHGSCKLPGDCRCQYGWQGLYCDKCIPHPGCVHGTCNEPWQCLCETNWGGQLCDKDLNYCGTHQPCLNRGTCSNTGPDKYQCSCPEGYSGPNCEIAEHACLSDPCHNRGSCKETSSGFECECSPGWTGPTCSTNIDDEFGLVPRGSGHHHHHH; human JAG1: MLLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFAADLGSQFELEILSMQNVNGELQNGNCCGGARNPGDRKCTRDECDTYFKVCLKEYQSRVTAGGPCSFGSGSTPVIGGNTFNLKASRGNDRNRIVLPFSFAWPRSYTLLVEAWDSSNDTVQPDSIIEKASHSGMINPSRQWQTLKQNTGVAHFEYQIRVTCDDYYYGFGCNKFCRPRDDFFGHYACDQNGNKTCMEGWMGPECNRAICRQGCSPKHGSCKLPGDCRCQYGWQGLYCDKCIPHPGCVHGICNEPWQCLCETNWGGQLCDKDLNYCGTHQPCLNGGTCSNTGPDKYQCSCPEGYSGPNCEIAEHACLSDPCHNRGSCKETSLGFECECSPGWTGPTCSTNIDDEFGLVPRGSGHHHHHH; mouse JAG2: MLLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFAADLGSYFELQLSALRNVNGELLSGACCDGDGRTTRAGGCGRDECDTYVRVCLKEYQAKVTPTGPCSYGYGATPVLGGNSFYLPPAGAAGDRARARSRTGGHQDPGLVVIPFQFAWPRSFTLIVEAWDWDNDTTPDEELLIERVSHAGMINPEDRWKSLHFSGHVAHLELQIRVRCDENYYSATCNKFCRPRNDFFGHYTCDQYGNKACMDGWMGKECKEAVCKQGCNLLHGGCTVPGECRCSYGWQGKFCDECVPYPGCVHGSCVEPWHCDCETNWGGLLCDKDLNYCGSHHPCVNGGTCINAEPDQYLCACPDGYLGKNCERAEHACASNPCANGGSCHEVPSGFECHCPSGWNGPTCALDIDEEFGLVPRGSGHHHHHH; human JAG2: MLLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFAADLGSYFELQLSALRNVNGELLSGACCDGDGRTTRAGGCGHDECDTYVRVCLKEYQAKVTPTGPCSYGHGATPVLGGNSFYLPPAGAAGDRARARARAGGDQDPGLVVIPFQFAWPRSFTLIVEAWDWDNDTTPNEELLIERVSHAGMINPEDRWKSLHFSGHVAHLELQIRVRCDENYYSATCNKFCRPRNDFFGHYTCDQYGNKACMDGWMGKECKEAVCKQGCNLLHGGCTVPGECRCSYGWQGRFCDECVPYPGCVHGSCVEPWQCNCETNWGGLLCDKDLNYCGSHHPCTNGGTCINAEPDQYRCTCPDGYSGRNCEKAEHACTSNPCANGGSCHEVPSGFECHCPSGWSGPTCALDIDEEFGLVPRGSGHHHHHH.

Anti-JAG1 antibody expression and Fab purification. Anti-JAG1 was expressed in CHO cells and purified from cell-conditioned media using MabSelect SuRe resin (GE Healthcare). After loading, resin was washed with five column volumes of: buffer A (25 mM Tris, pH 7.5, 150 mM NaCl and 5 mM EDTA), buffer B (25 mM Tris, pH 7.5, 150 mM NaCl, 5 mM EDTA and 0.1% (v/v) Triton X-114), buffer C (400 mM potassium phosphate, 5 mM EDTA and 0.2% (v/v) Tween 20), and again with buffer A. The antibody was eluted with buffer D (50 mM sodium acetate pH 3.0, 50 mM NaCl), neutralized by adding 1.5 M Tris, pH 9.0 (to pH 7.0), and purified over a Superdex 200 column (GE Healthcare) in PBS, pH 7.4, 150 mM NaCl. Peak fractions were incubated with lysyl endopeptidase (Wako Chemicals, Inc.) at 37 °C for 1 h to generate the anti-JAG1 Fab and cleavage was stopped by addition of sodium acetate (pH 3.0, 250 mM final). Finally, the Fab was purified over an SP Sepharose Fast Flow column (GE Healthcare) using a 0–30% (w/v) NaCl gradient and molecular mass was confirmed by liquid chromatography–time-of-flight mass spectrometry (LC–MS TOF).

JAG1 DSL–EGF1–4 expression and purification. Human JAG1 (V187–D377; ECD) was cloned, expressed and purified as described above for other JAG1 and JAG2 proteins. After purification over a Superdex 75 column in buffer P3 (defined above), peak fractions were incubated with TEV protease to remove the His tag. Cleaved JAG1 was collected in the flow-through from a Ni-NTA column and repurified over a Superdex 75 column in the same buffer.

Crystallization and structure determination. The JAG1-ECD–anti-JAG1.b70 Fab complex was prepared by adding a molar excess of JAG1 ECD and purified over a Superdex S200 column (GE Healthcare) in buffer P3. Peak fractions containing the complex were concentrated to 6 mg ml^{-1} and crystallized by sitting drop vapour diffusion in a 1:1 ratio with 100 mM CHES, pH 9.5, 20% PEG8000 at 19 °C. Crystals were cryoprotected with 15% (v/v) glycerol before flash freezing in liquid nitrogen and maintained at 100 K during data collection. Native data were collected at the Stanford Synchrotron Radiation Lightsource (SSRL 12.2) using a PILATUS6M PAD detector. Diffraction data were integrated and scaled using autoPROC (Global Phasing)[34] to 2.56 Å and, when required, further processed using the CCP4 package[35]. Available JAG1 ECD (PDB 2VJ2) and Fab (PDB 2R0L) structural models were used after trimming CDRs to poly-Ala for the initial molecular replacement solution using the Phenix software[36]. Reiterative building using COOT[37] and refinement in Buster (Global Phasing)[38] were used to arrive at the final model JAG1–anti-JAG1.b70 model. EGF4 was present in the crystallized JAG1 construct but not modelled, as only diffuse electron density was observed for this domain. Two independent JAG1–anti-JAG1.b70 complexes are present in the asymmetric unit of essentially identical structure. Geometry was assessed using PROCHECK[39] and MolProbity[40] and structural figures were prepared with the PyMol software[41].

Generation of Scgb1a1-CreERT2^{GNE} mice. The construct for targeting the CreERT2 recombinase into the C57BL/6 *Scgb1a1* locus in embryonic stem (ES) cells was made using a combination of recombineering and standard molecular cloning techniques. In brief, a cassette (CreERT2 SV40 pA, and frt-PGK-em7-Neo-BGHpA-frt) flanked by short homologies to the mouse *Scgb1a1* gene was used to modify an *Scgb1a1* C57BL/6J bacterial artificial chromosome (BAC) (RP23-234B14) by recombineering. The CreERT2 cDNA cassette was inserted at the endogenous ATG and the remainders of the Scgb1a1 exon 1 plus the beginning of intron 1 were deleted. The targeted region in the BAC was then retrieved into pBlight-TK along with flanking genomic *Scgb1a1* sequences as homology arms for ES cell targeting. Specifically, the 2939–base pair (bp) 5′ homology arm corresponds to NCBI37/mm9 chr19:9,162,392-9,165,330 (reverse strand) and the 2660-bp 3′ homology arm corresponds to chr19:9,159,677-9,162,336 (reverse strand). The final vector was confirmed by DNA sequencing. The vector was linearized with Not I, and C57BL/6N C2 ES cells were targeted using standard methods (G418-positive and gancyclovir-negative selection). Positive clones were identified using polymerase chain reaction (PCR) and TaqMan analysis and confirmed by sequencing of the modified locus. Correctly targeted ES cells were transfected with an FLPe plasmid to remove Neo, and ES cells were then injected into blastocysts using standard techniques. Germline transmission was obtained after crossing resulting chimaeras with C57BL/6N females. Founders were determined by long PCR sequencing and screening for club cell specific recombination after crossing to the Rosa26-lsl-tdTomato reported mouse (Jax stock 007914). Reporter expression was not observed in the airways of adult mice without Tamoxifen injection.

Genotyping of Scgb1a1-CreERT2^{GNE} mice. For genotyping three primers were used, forward1-(F1): 5′-TCTCCTAAGTGGAGCGCAATC-3′, forward2-(F2): 5′-GCATCTGTACAGCATGAAGTGC-3′ and reverse-(R): 5′-GACGCAATGCTTCTGAGAGTC-3′. PCR amplification yielded a 295-bp product for the wild-type allele (F1+R primers) and a 646-bp product for the knock-in allele (F2+R primers).

Mice. Animal studies were conducted in accordance with the Guide for the Care and Use of Laboratory Animals, published by the National Academy Press (2006). Female BALB/c and C57BL/6 mice were obtained from Jackson Laboratories or Charles River Laboratories. Rosa26-lsl-tdTomato mice (stock 007914) were obtained from The Jackson Laboratory and maintained on a C57BL/6 background. All mice were housed under specific pathogen-free (SPF) conditions and used at 8–12 weeks of age. Investigators performing mouse experiments were not blinded. The Genentech Institutional Animal Care and Use Committee (IACUC) approved all animal studies. Mice were injected intraperitonealy with blocking antibodies diluted in 200 μl of PBS at the following concentrations: anti-JAG1.b70 at 15 mg kg^{-1}, anti-JAG2.b33 at 15 mg kg^{-1}, anti-NRR1 at 10 mg kg^{-1}, anti-NRR2 at 30 mg kg^{-1}, anti-ragweed (non-targeting control antibody) at concentrations equal to the maximum dose of blocking antibodies. Anti-ragweed was also supplemented to achieve equal dose of total antibody injected in each study.

For the lineage-tracing studies, both male and female Scgb1a1-CreERT2^{GNE}/Rosa26-lsl-tdTomato were used. Compound transgenic mice were induced with

four doses of 200 mg kg^{-1} tamoxifen (T5648, Sigma) diluted in sesame seed oil and were subsequently allowed 1 week to recover. BrdU was added to drinking water supplemented with 5% sucrose at a concentration of 1 mg ml^{-1} and was renewed every three days.

For intranasal administration of antibodies, each of the anti-JAG1.b70 and anti-JAG2.b33 were diluted at a concentration of 4 mg ml^{-1} in saline and anti-ragweed control was diluted to a concentration of 8 mg ml^{-1}. After anaesthesia with avertin, 62.5 μl of antibodies was instilled over two doses, delivered as two 31.25-μl doses into each nostril. Mice were allowed to recover and were analysed five days later.

Immunohistological staining. Lungs were cleared of blood by right ventricular perfusion with saline solution containing 2 U ml^{-1} heparin, inflated with 4% paraformaldehyde (PFA) in PBS and submerged in Zfix overnight. Individual lobes were separated and embedded in paraffin and subsequently sectioned at 5 μm. Sections were boiled for 15 min in target retrieval solution (DAKO-S1700) in a pressure cooker resulting in rehydration and antigen retrieval. Sections were then permeabilized with 0.2% Triton X-100 in PBS for 45 min and blocked for 1 h with 5% FBS and 2% BSA in PBS. For BrdU staining, paraffin was removed from sections by treating with xylenes, and the sections were rehydrated through a gradient of ethanol before a 1-h incubation in 2 N HCl. After washing in PBS, sections were incubated with 0.05% Trypsin/EDTA for 30 min and then washed with PBS before immunofluorescence staining as above. Primary antibodies were incubated in blocking buffer (5% FBS, 2% BSA in PBS) for 2 h at room temperature. The following primary antibodies were used: anti-CC10 (1:1,000; sc9772, Santa Cruz), anti-acetylated-α-tubulin (1:200; sc23950, Santa Cruz), anti-FOXJ1 (1:200; 14-9965-82, eBioscience), anti-BrdU (1:200; 347580, BD), anti-tdTomato (1:1,000; 600-401-379, Rockland). Secondary antibodies were incubated in blocking buffer for 1–2 h at room temperature. The following secondary antibodies were used: anti-mouse Alexa Fluor 488 (1:1,000; A21202, Invitrogen), anti-rabbit Alexa Fluor 488 (1:1,000; A-21206, Invitrogen), anti-goat Alexa Fluor 488 (1:1,000; A-11055; Invitrogen), anti-mouse Alexa Fluor 555 (1:1,000; A31570, Invitrogen), anti-rabbit Alexa Fluor 555 (1:1,000; A-31572, Invitrogen), anti-goat Alexa Fluor 555 (1:1,000; A-21432; Invitrogen), anti-mouse Alexa Fluor 647 (1:1,000; A31571, Invitrogen), anti-rabbit Alexa Fluor 647 (1:1,000; A-31573, Invitrogen), anti-goat Alexa Fluor 647 (1:1,000; A-21417; Invitrogen). Nuclei were stained with DAPI. Slides were imaged using an Olympus BX-61 upright wide field microscope equipped with the following objectives: 10× UPlanS APO 0.4 numerical aperture (NA), 20× UPlanS APO 0.75 NA, 40× UPlanS APO 0.9 NA and 60× UPlan FLN 0.9 NA, as well as the following filters: DAPI (ex. 387/11, em. 447/60), FITC (ex. 482/35, em. 536/40), Cy3 (ex. 531/40, em. 593/40), Cy5 (ex. 628/40, em. 692/40). Images were obtained using SlideBook software (3i - Intelligent Imaging Innovations) and were pseudo-coloured and edited using Photoshop CS6 (Adobe).

Immunohistochemistry staining was performed on 4-μm thick formalin-fixed, paraffin-embedded tissue sections mounted on glass slides as previously described[33]. In brief, primary antibodies against JAG1, polyclonal (sc-6011, Santa Cruz Biotechnology), Notch-1, clone D1E11 (3608, Cell Signaling Technologies), Notch-2, clone D76A6 (5732 s, Cell Signaling Technologies) and Hes-1, clone NM1 (D134-3, MBL International) were used at 0.7 μg ml^{-1}, 5 μg ml^{-1}, 8 μg $ml^{-1}$and 1 μg ml^{-1} respectively. NOTCH1, NOTCH2 and JAG1 staining was carried out on the Ventana Discovery XT automated platform (Ventana Medical Systems). Sections were treated with Cell Conditioner 1, standard time. Specifically bound primary antibody was detected by incubating sections in OmniMap anti-rabbit-HRP (Ventana Medical Systems) for NOTCH1, NOTCH2 and OmniMap anti-goat-HRP (Ventana Medical Systems) for JAG1 followed by ChromoMap DAB (Ventana Medical Systems). HES1 staining was performed on the DAKO autostainer, using Target pH6 (Dako) antigen retrieval. Detection used donkey anti-rat biotinylated secondary (Jackson ImmunoResearch Laboratories), followed by streptavidin-HRP with TSA enhancement (PerkinElmer) and DAB visualization (Pierce). The sections were counterstained with haematoxylin and dehydrated. For the Alcian blue/PAS staining, tissues were fixed 24 h at ambient temperature in 10% neutral buffered formalin (VWR) then processed and embedded using a Tissue-Tek VIP processor (Sakura). Four-micrometre sections were mounted on Superfrost Plus glass slides (Richard-Allan) and dried for 30 min at 60 °C, then dewaxed and rehydrated before staining. Alcian blue/PAS staining was performed using an Artisan automated stainer and staining reagents according to the manufacturer's instructions (AR16911-2, DAKO). Slides were then air-dried, cleared with xylene, and mounted with a synthetic mounting medium (Tissue-Tek Glas, Sakura).

Standard (morphology) transmission electron microscopy. For standard transmission electron microscopy the lung tissues were fixed in 1/2 Karnovsky's fixative (2% PFA, 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer, pH 7.2). The samples were post-fixed in 1% aqueous osmium tetroxide for 2 h, stained with 0.5% uranyl acetate for 1 h and then dehydrated through a series of ethanol (50%, 70%, 90%, 100%) followed by two propylene oxide washes. Samples were embedded in Eponate 12 (Ted Pella). Curing of the samples was at 65 °C for 2 days. Semithin (300 nm) and ultrathin (80 nm) sections were obtained with an Ultracut microtome (Leica). The semithin sections were stained with Toluidine Blue and examined by bright field microscopy to identify tissue areas with terminal bronchioles. Then parallel ultrathin sections were prepared, counter stained with 0.2% lead citrate and examined in a JEOL JEM-1400 transmission electron microscope at 80 kV. Digital images were captured with a GATAN Ultrascan 1000 CCD camera.

Immunogold electron microscopy. For immuno-electron micoscopy studies, lungs were fixed in 4% PFA in 0.1 M phosphate buffer (pH 7.2) for several days and stored at 4 °C. Tissue samples were cut into 1-mm pieces, washed in PBS and quenched for 5 min in 0.15% glycine in PBS before being rinsed in water. Samples were then dehydrated with an ascending series of dimethylformamide in water followed by two 100% dimethylformamide steps; each step for 15 min at 4 °C. The tissues were finally infiltrated with LR White resin (London Resin Company) and cured at 55 °C for 2 days. Semithin (300 nm) and ultrathin (80 nm) sections were obtained with an Ultracut microtome (Leica). The semithin sections were transferred to glass slides, stained with Toluidine Blue and examined by bright field microscopy to identify areas containing terminal bronchioles. Light microscopy images were acquired at 1,000 × using a Zeiss Axioplan microscope and a Zeiss Plan Apochromat objective (100 ×, 1.4 NA, oil immersion). Ultrathin (80 nm) sections of areas containing terminal bronchioles were then transferred to transmission electron microscopy grids and used for immunogold labelling. Grids were simultaneously incubated with a (mouse) monoclonal antibody for acetylated-α-tubulin (Abcam, ab24610, dilution 1:25) and a (goat) polyclonal antibody for CC10 (Santa Cruz, sc9772, dilution 1:100). Secondary antibodies were (donkey)-anti-mouse-12nm and (donkey)-anti-goat-18 nm antibody–gold conjugates (Jackson ImmunoResearch) diluted at 1:20 and 1:10, respectively. Labelled sections were counterstained with 0.5% uranyl acetate in water for 5 min at room temperature and examined in a JEOL JEM-1400 transmission electron microscope at 80 kV. Digital images were captured with a GATAN Ultrascan 1000 CCD camera at magnifications from 500–50,000×. Specificity of the labelling was determined by confirming absence of labelling in a negative control (no primary antibodies, but all secondary antibodies were used), absence of labelling in unrelated areas and confirmation of expected labelling over cilia (anti-acetylated tubulin antibody) and secretory vesicles (anti-CC10 antibody), respectively, in single-labelling experiments.

Scanning electron microscopy. For scanning electron microscopy, 1-mm thick sections through the lobes of the lungs were fixed in 1/2 Karnovsky's fixative (see earlier). The samples were then post-fixed in 1% aqueous osmium tetroxide, stained with 1% uranyl acetate for 2 h and then dehydrated through a series of ethanol (50%, 70%, 90%, 95%, 100%) followed by three changes in 100% hexamethyldisilazane. Finally, the samples were mounted on scanning electron microscopy stubs, air dried and coated with 5-nm palladium-gold using an EMS150R ES sputter coater (Electron Microscopy Sciences). The samples were imaged with an FEI XL30 ESEM in secondary electron mode at 5 kV and 15 mm working distance. Images were acquired at magnifications from 500× to 5,000×.

Whole-mount staining. Staining was performed as previously described[42]. Following perfusion of the lung to remove blood, lungs were inflated and fixed in 4% PFA overnight. The right caudal was removed and placed in 1% Triton X-100 in PBS until the tissue sank to the bottom of the tube. Whole-mount lobes were stained with anti-RFP (1:1,000; 600-401-379, Rockland) for 72 h in 4 ml blocking buffer (5% FBS, 2% BSA in PBS) containing 0.2% Triton X-100 at 4 °C and were subsequently washed in a large volume of 0.2% Triton X-100 in PBS for 6–8 h. Whole-mount lobes were subsequently stained with anti-rabbit Alexa Fluor 555 (1:1,000; A-31572, Invitrogen) secondary antibody overnight in 4 ml blocking buffer containing 0.2% Triton X-100 at 4 °C and were subsequently washed in a large volume of 0.2% Triton X-100 in PBS for 6–8 h and were subsequently stored in PBS until clearing.

Optical clearing of whole-mounts. For optical clearing, lobes were dehydrated through a gradient of tetrahydrofurane: 50% for 30 min, 70% for 30 min, 80% for 30 min, and three times 100% for 30 min each. Dehydrated lobes were subsequently incubated in dimethyl ether for 20 min followed by 30 min in dibenzyl ether to obtain an optically cleared sample. Imaging was performed using the LaVision Ultramicroscope (La Vision BioTec GmbH) composed of an Olympus MVX10 stereomicroscope equipped with a 2 × 0.3 NA SFD-PLAPO air objective and a sCMOS pco.edge camera. Images were acquired using Imspecter software (La Vision BioTec GmbH) and were subsequently analysed with Fiji.

RNA extraction and qPCR. For RNA extraction, half a lung was lysed and homogenized in 4 ml buffer RLT with 2-mercaptoethanol. Lysates (350 μl each) were run through a Qiashredder column and RNA isolation was completed with the Qiagen RNeasy kit as per the manufacturer's instructions with DNase digestion. Eluted RNA was resuspended in 60 μl total volume of RNase-free water. cDNA was synthesized used the ABI High Capacity cDNA Reverse Transcription Kit, using

200 ng RNA in 20 μl total reaction volume as per the manufacturer's instructions. cDNAs were preamplified and prepared for quantitative Real Time PCR as per the Preamp Protocol: Fluidigm Specific Target Amplification. All real time PCR reactions were run on the Fluidigm platform with the following TaqMan assays: *Gapdh* (Mm99999915_g1), *Hes1* (Mm01342805_m1), *Hey1* (Mm00468865_m1), *Hey2* (Mm00469280_m1), *Foxj1* (Mm01267279_m1), *Scgb1a1* (Mm00442046_m1), *Muc5b* (Mm00466391_m1) and *Muc5ac* (Mm01276718_m1).

RNAscope *in situ* hybridization. RNA *in situ* hybridization was performed by Advanced Cell Diagnostics, Inc. for mouse *Notch1* (404641-C3), *Notch2* (425161-C1), *Jag1* (412831-C1), *Jag2* (417511-C1), *Foxj1* (317091-C2) and *Scgb1a1* (420351-C3) mRNA was performed manually using RNAscope Multiplex Fluorescent Reagent Kit (320850) according to the manufacturer's instructions. In brief, 5-μm formalin-fixed, paraffin-embedded tissue sections were pre-treated with heat and protease before hybridization with the target oligonucleotide probes. Preamplifier, amplifier and alkaline-phosphatase-labelled oligonucleotides were then hybridized sequentially, coupled with a fluorescent conjugate. Each sample was quality controlled for RNA integrity with an RNAscope probe specific to PolR2A/PPIB/UBC RNA (320881) and for background with a probe specific to bacterial *dapB* RNA (320881). Specific RNA staining signal for each of *Notch1*, *Notch2*, *Jag1*, *Jag2* was identified as green, *Foxj1* RNA was identified as red and *Scgb1a1* RNA was identified as white punctate dots. Samples were counterstained with DAPI.

Analysis of clonal expansion of lineage traced cells. For the quantification of the clonal expansion of lineage traced club cells Scgb1a1-CreERT2^{GNE}/Rosa26-lsl-tdTomato compound transgenic mice were induced with four doses of 200 mg kg^{-1} tamoxifen (T5648, Sigma) diluted in sesame seed oil and were subsequently allowed 1 week to recover. Four mice were subsequently treated with a single dose of either anti-ragweed non-targeting isotype control antibody at 30 mg kg^{-1} or anti-JAG1.b70 at 15 mg kg^{-1} plus anti-JAG2.b33 at 15 mg kg^{-1}. Lungs were sectioned and stained by immunofluorescence for the lineage-tracing marker tdTomato, CC10 to mark club cells and acetylated-α-tubulin or FOXJ1 to mark ciliated cells, 13 weeks after treatment. Clones were defined as single cells, small clusters of 2–4 cells or clusters of more than 5 cells. Quantifications were done by eye and were calculated as the number of one of the three types of clones over the total amount of clones.

Single-cell isolation and video capture. A single-cell suspension was prepared from treated mouse tracheas as previously described[43]. In brief, tracheas were resected from the bronchial bifurcation to just distal to larynx, cleared from adherent tissue and incubated in 0.2% (w/v) pronase (Roche Applied Science) in Ham's F12 medium (Life Technologies) containing 1% penicillin/streptomycin (Life Technologies) overnight at 4 °C. Imaging was performed using a 20× S Plan Fluor objective (0.45 NA; Nikon) on a Nikon Ti-E perfect focus inverted microscope equipped with Live Cell environmental chamber (Pathology Devices), Neo sCMOS camera (Andor, Oxford Instruments), and controlled by NIS-Elements software (Nikon). Red fluorescence and phase channels were acquired simultaneously at 90 frames per second using Fast Timelapse mode in NIS Elements. Images are displayed in greyscale, bright white cells are expressing tdTomato. Time-lapse acquisitions were analysed in NIS-Elements, exported as AVI files and playback in real time.

Ovalbumin-induced goblet cell metaplasia model. Seven-to-eight-week-old C57BL/6 mice were first sensitized by i.p. injection of 50 μg ovalbumin in 2 mg alum. Then 35 days later, mice were challenged for 7 consecutive days with 1% ovalbumin aerosol for 30 min in a chamber. For the prevention study, blocking antibodies were administered i.p. on days 36 and 39. The concentrations of the antibodies were as follows: anti-ragweed control non-targeting antibody 30 mg kg^{-1}, anti-JAG1.b70 15 mg kg^{-1}, anti-JAG2.b33 15 mg kg^{-1}, anti-NRR1 10 mg kg^{-1}, anti-NRR2 20 mg kg^{-1}. Lungs were collected 24 h after the last challenge for analyses of inflammatory infiltrate and Alcian blue and PAS staining. For the intervention study, mice were challenged with inhaled ovalbumin on days 42 and 45 and treated with blocking antibodies on the same days. Lungs were collected on day 48 for analyses of inflammatory infiltrate and Alcian blue and PAS staining. At both endpoints, bronchoalveolar lavage fluid was collected and control, anti-JAG1.b70 and anti-JAG1.b70 plus anti-JAG2.b33 were analysed for number of immune cell populations and cytokine concentrations by Luminex. Blood serum was also collected and cytokine concentrations were determined by Luminex.

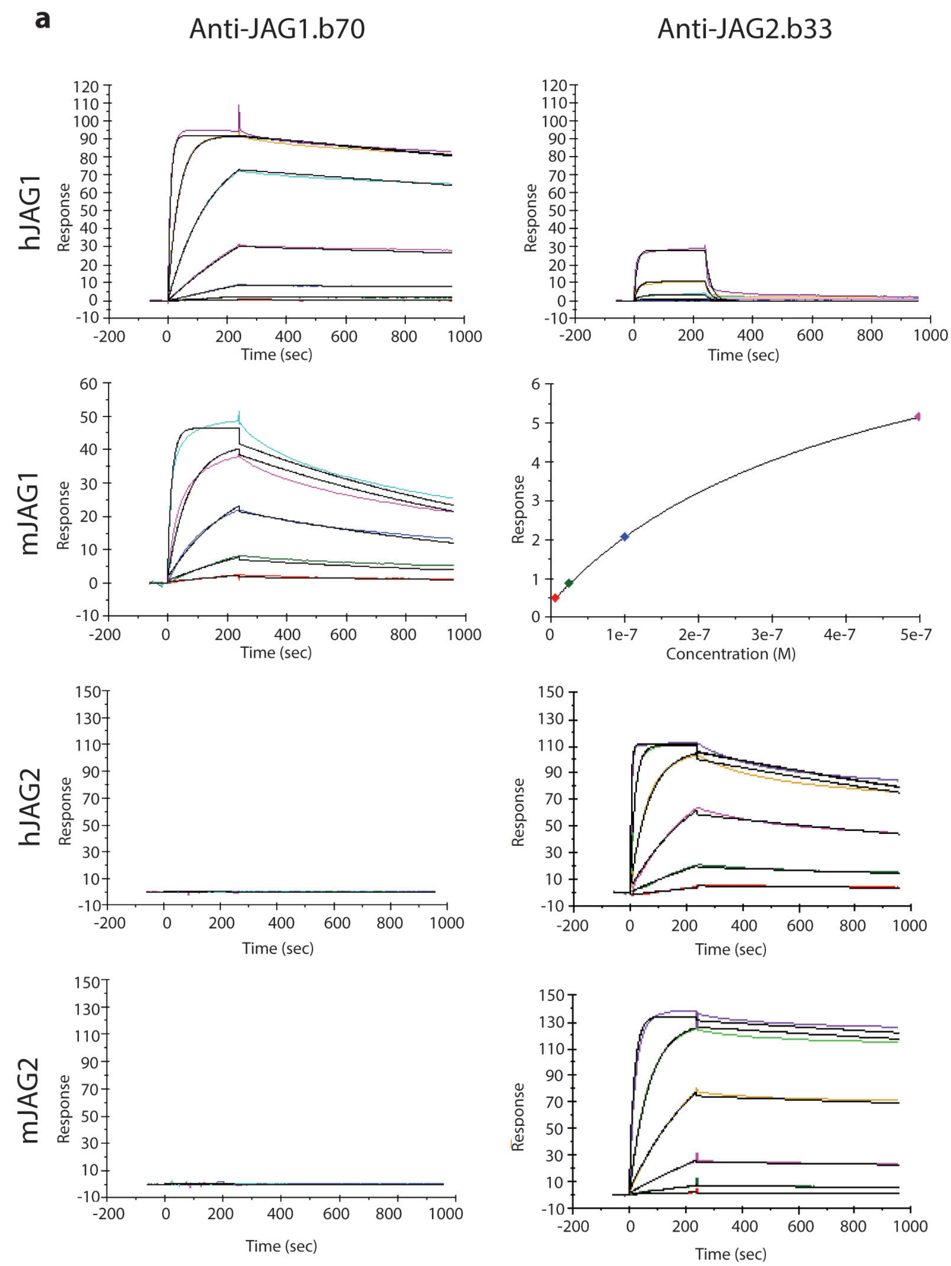
Histopathology analysis. Lung tissues were fixed for 24 h at ambient temperature in 10% neutral buffered formalin (VWR), then processed and embedded using a Tissue-Tek VIP processor (Sakura). Sections (4 μm) were mounted on Superfrost Plus glass slides (Richard-Allan) and stained with haematoxylin and eosin using a Leica Autostainer XL or with Alcian blue and PAS using an Artisan automated stainer (DAKO) according to the manufacturer's instructions. Slides were then air-dried, cleared with xylene, and mounted with a synthetic mounting medium (Tissue-Tek Glas). Inflammation severity was visually scored in a blinded fashion on haematoxylin-and-eosin-stained slides using a subjective semi-quantitative five-point scale, in which 0 = normal lung (no inflammatory infiltrate), 1 = minimal disease (infrequent sparsely scattered inflammatory cells), 2 = mild (light perivascular/peribronchiolar involvement), 3 = moderate (many vessels and airways affected by substantial numbers of inflammatory cells), and 4 = severe (generalized accumulations of perivascular/peribronchiolar inflammatory cells with frequent circumferential and/or bridging infiltrates). Goblet cell area was quantified on Alcian blue/PAS-stained slides imaged with a Nanozoomer 2.0-HT automated slide scanning platform (Hamamatsu) at 200× final magnification. Slide images were analysed in the Matlab software package (version R2012b by Mathworks) as 24-bit RGB images. Regions of interest (ROIs) corresponding to individual profiles of medium and large airways were defined using RGB thresholding and simple morphological and shape-based filtering. Airway ROIs were subject to manual curation to remove false positives corresponding to vessels and other non-airway regions. Epithelial area within each airway ROI was defined using a similar approach. RGB thresholding was used to identify Alcian blue/PAS-positive epithelial area and the data was normalized to either cumulative airway ROI epithelial area or cumulative airway ROI perimeter. For immunofluorescence quantifications (Fig. 2c, f, h) standard morphological operations were used to identify airways with characteristically dense DAPI staining surrounded by empty areas that also contained CC10 or FOXJ1 staining. The total number of cells in the airway were counted and scored as positive for CC10 or FOXJ1, KI67 or BrdU respectively. For the quantification of Fig. 3d, tdTomato cells on the periphery of each airway were identified using an algorithm based on radial symmetry[44]. Each tdTomato-positive cell was then scored as positive or negative for BrdU staining. Quantifications of double-positive cells (Fig. 3c and Extended Data Fig. 6e) were done by manual counting on whole slide images. The raw data as well as the averages and standard deviations for these quantifications can be found in Supplementary Table 1.

Cytokine concentration measurement. Cytokine concentrations were measured by using the Biorad Bio-Plex 200 system and the Bio-Plex Manager Software v6.0. The Bio-Plex Pro Mouse Cytokine 9-Plex Assay (MD000000EL) and 23-plex Assay (M60009RDPD) were used. Concentrations of cytokines previously reported as relevant to goblet cell metaplasia are shown.

Statistical analysis. For mouse studies, the phenotypes described were clear, marked (near-complete conversion) and reproducible. We used 5–7 mice per group when generating data for statistical analysis and quantification. We randomized animals between the groups based on sex and weight, to normalize the distribution of these two parameters. We conducted an unpaired *t*-test to ascertain statistical significance, using the GraphPad Prism software. No statistical method was used to predetermine sample size. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. No sample or animal was excluded from analyses. The investigators were not blinded for any of the studies.

31. Lee, C. V., Sidhu, S. S. & Fuh, G. Bivalent antibody phage display mimics natural immunoglobulin. *J. Immunol. Methods* **284,** 119–132 (2004).
32. Liang, W. C. *et al.* Function blocking antibodies to neuropilin-1 generated from a designed human synthetic antibody phage library. *J. Mol. Biol.* **366,** 815–829 (2007).
33. Huntzicker, E. G. *et al.* Differential effects of targeting Notch receptors in a mouse model of liver cancer. *Hepatology* **61,** 942–952 (2015).
34. Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D* **67,** 293–302 (2011).
35. Collaborative Computational Project, Number 4. The *CCP4* suite: Programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
36. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
37. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
38. Bricogne G. *et al.* BUSTER version 2.11.2 (2011).
39. Laskowski, R. A., Moss, D. S. & Thornton, J. M. Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231,** 1049–1067 (1993).
40. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
41. Schrödinger, L. The PyMOL Molecular Graphics System version 1.2r3pre edn.
42. Ertürk, A., Lafkas, D. & Chalouni, C. Imaging cleared intact biological systems at a cellular level by 3DISCO. *J. Vis. Exp.* **89,** 51382 (2014).
43. You, Y. & Brody, S. L. Culture and differentiation of mouse tracheal epithelial cells. *Methods Mol. Biol.* **945,** 123–143 (2012).
44. Veta, M. *et al.* Automatic nuclei segmentation in H&E stained breast cancer histopathology images. *PLoS ONE* **8,** e70221 (2013).

a

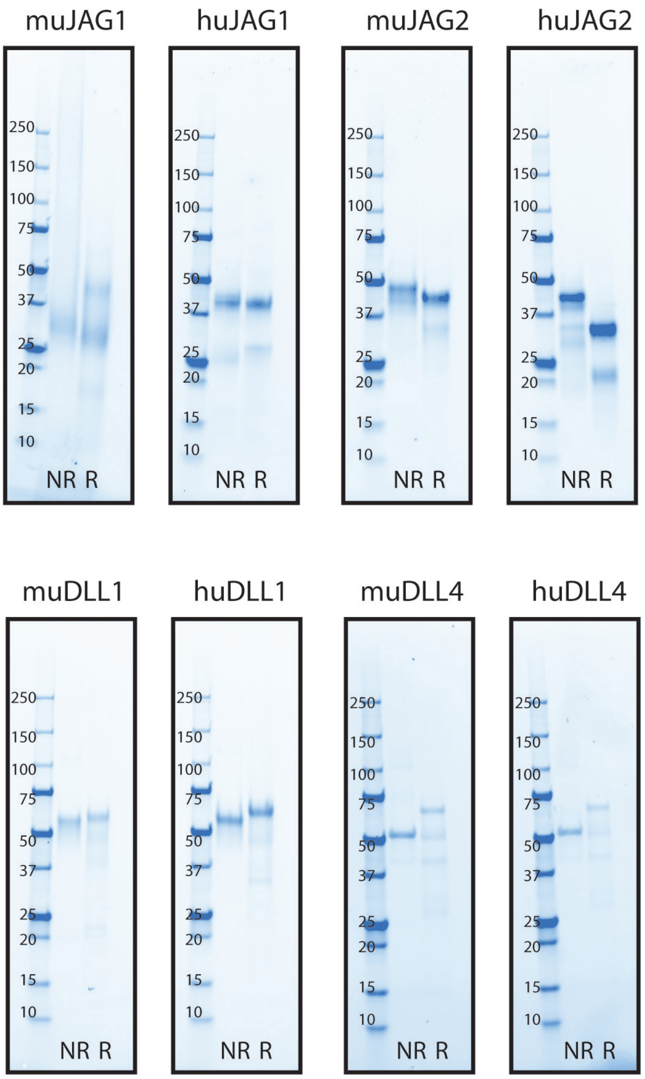


b

	Anti-JAG1.b70 k_{on} ($M^{-1}s^{-1}$)	Anti-JAG1.b70 k_{off} (s^{-1})	Anti-JAG1.b70 K_d (M)	Anti-JAG1.b70 χ^2(RU^2)	Anti-JAG2.b33 k_{on} ($M^{-1}s^{-1}$)	Anti-JAG2.b33 k_{off} (s^{-1})	Anti-JAG2.b33 K_d (M)	Anti-JAG2.b33 χ^2(RU^2)
hJAG1	$(2.29\pm0.02)\times10^{5}$	$(1.8\pm0.1)\times10^{-4}$	$(7.9\pm0.4)\times10^{-10}$	3 ± 1	$(1.5\pm0.6)\times10^{5}$	0.1 ± 0.04	$(7\pm1)\times10^{-7}$	0.26 ± 0.02
mJAG1*	$(1.3\pm0.6)\times10^{5}$	$(7.8\pm0.4)\times10^{-4}$	$(6.00\pm0.06)\times10^{-9}$	2.0 ± 0.5	steady state measurement		$(7\pm3)\times10^{-7}$	$(1.40\pm0.2)\times10^{-3}$
hJAG2	no binding				$(1.91\pm0.06)\times10^{6}$	$(4.02\pm0.07)\times10^{-4}$	$(2.10\pm0.06)\times10^{-10}$	2.7 ± 0.8
mJAG2	no binding				$(4.8\pm0.1)\times10^{5}$	$(9.8\pm0.4)\times10^{-5}$	$(2.0\pm0.1)\times10^{-10}$	9 ± 2
hDLL1	no binding				no binding			
mDLL1*	no binding				no binding			
hDLL4	no binding				no binding			
mDLL4	no binding				no binding			

Extended Data Figure 1 | Surface plasmon resonance affinity measurements of anti-JAG1.b70 and anti-JAG2.b33 binding. a, Surface plasmon resonance (SPR) was used to determine anti-JAG1.b70 and anti-JAG2.b33 binding affinities to purified human (h) or mouse (m) JAG1 and JAG2 antigens. Representative curves from one assay run with three technical replicates are shown. At least two additional assays have been performed with binding to human JAG1 and JAG2 antigens, yielding consistent results. **b**, SPR binding constants. For human JAG1, human and mouse JAG2, human DLL1, and human and mouse DLL4, antibodies were coated onto a CM5 biosensor chip and the ligand was subsequently added for binding assessment. *By contrast, because mouse JAG1 and DLL1 showed some background binding to the empty flow cell, the antigens were coated directly onto the CM5 biosensor chip, and purified antibodies in Fab fragment format were subsequently added for binding assessment; steady-state measurements were used for the low-affinity binding of anti-JAG2.b33 to mouse JAG1. Data are mean ± s.d. of three technical replicates. See Methods for details.

a



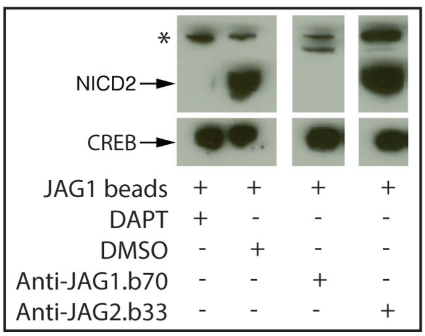
b

	huJAG1 - anti-JAG1.b70
Data collection	
Space group	*P*2₁
Cell dimensions	
a, *b*, *c* (Å)	53.26, 80.98, 155.62
a, *b*, *g* (Υ)	90, 94.39, 90
Resolution (Å)	80.98 – 2.56 (2.86 – 2.56) *
$CC_{1/2} > 0.5$ (Å)	2.56
R_{merge}	0.13 (0.57)
I/*sI*	6.4 (1.8)
Completeness (%)	96.6 (97.9)
Redundancy	3.3 (3.3)
Number of unique observations	41,335 (11,823)
Refinement	
Resolution (Å)	71.79 – 2.56
No. reflections	41,207
R_{work}/R_{free}	22.0 / 25.9
No. atoms#	
Protein	8746
Water/glycerol	170
B-factors	
Protein	41.58
Water/glycerol	33.62
R.m.s deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.04

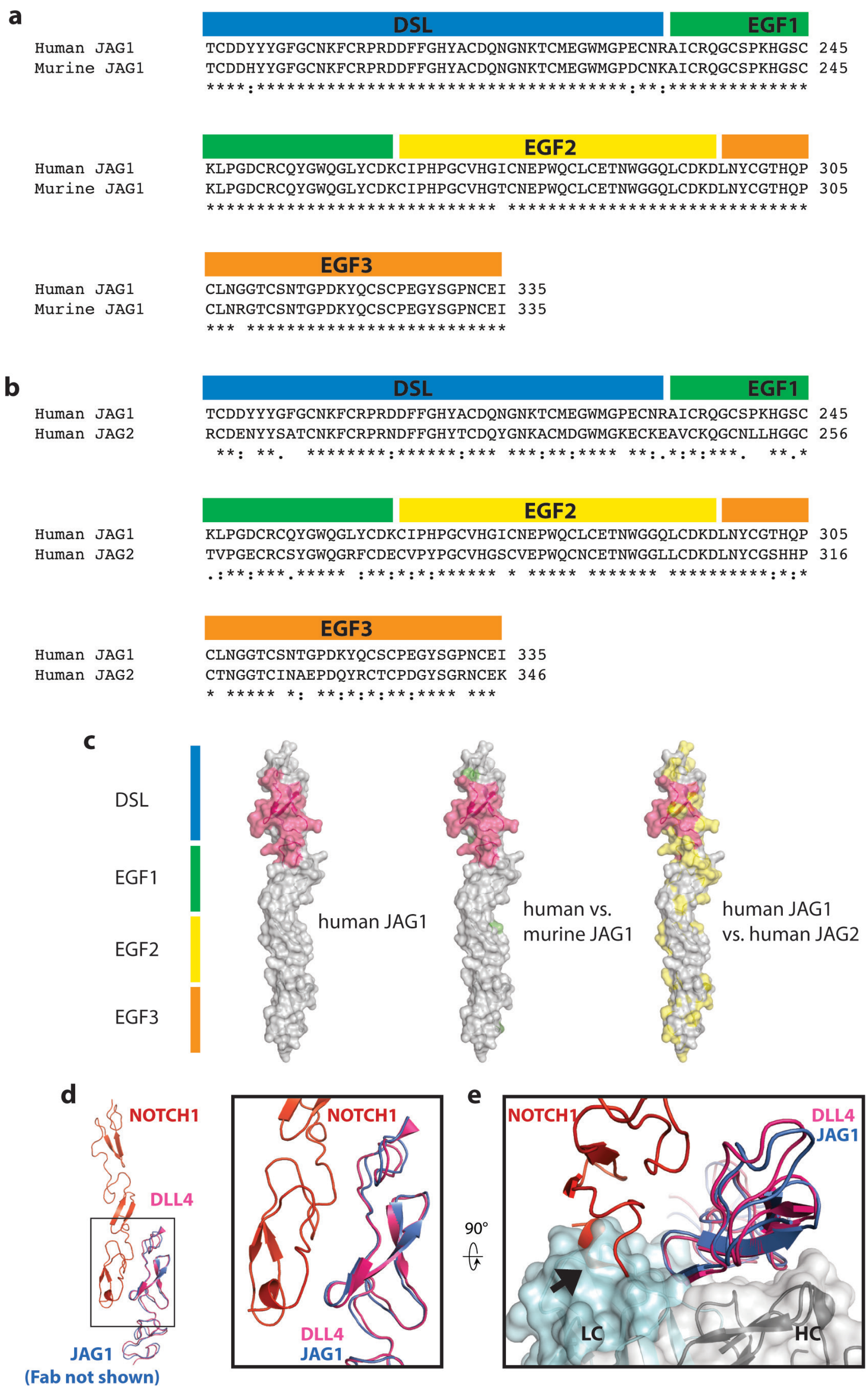
*Highest resolution shell is shown in parenthesis.
#sugars on huJAG1 were not modeled

Ramachandran statistics as defined by Molprobity:
favored 94.6% (1076), allowed 4.6% (53), outlier 0.8% (9)

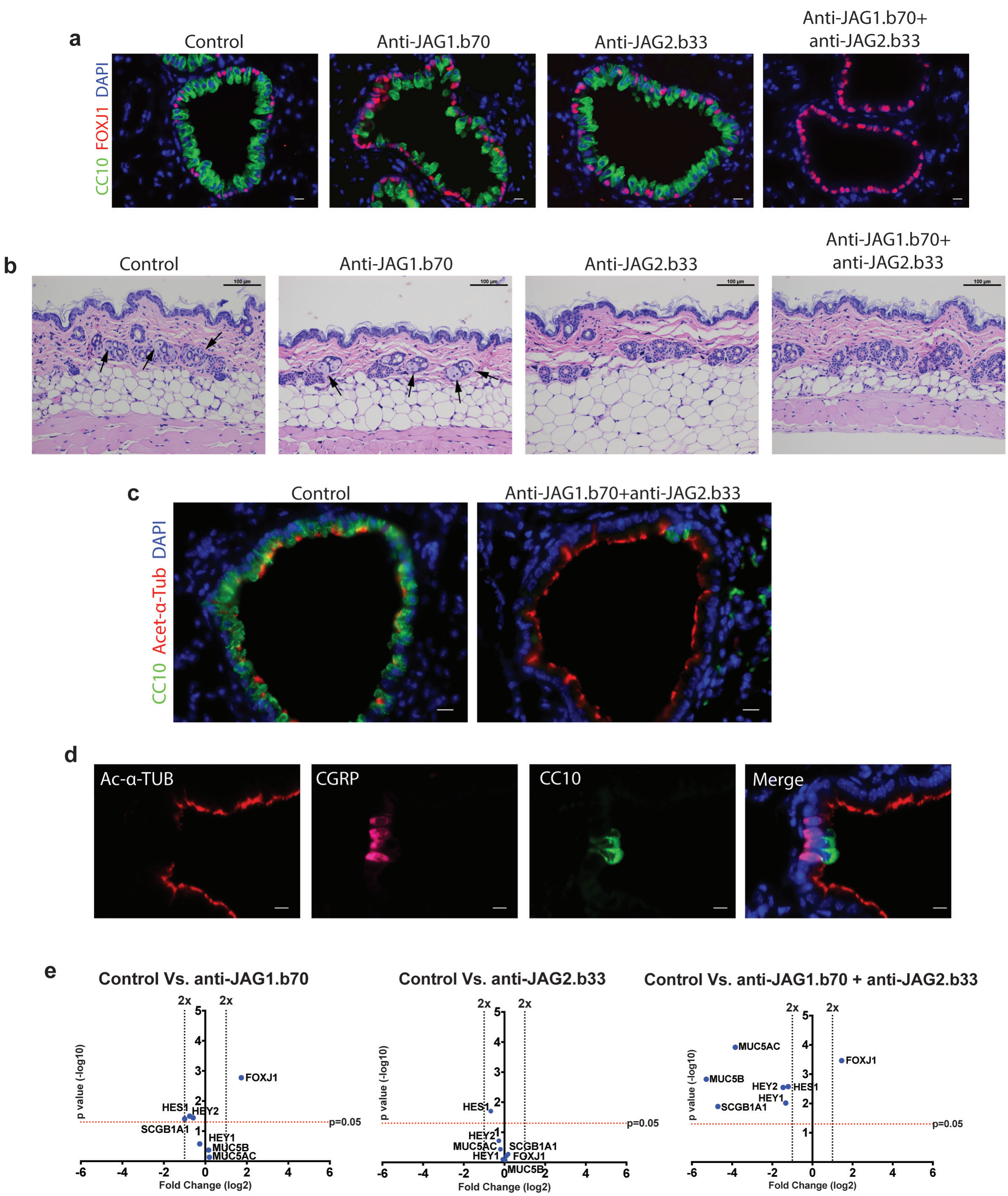
c



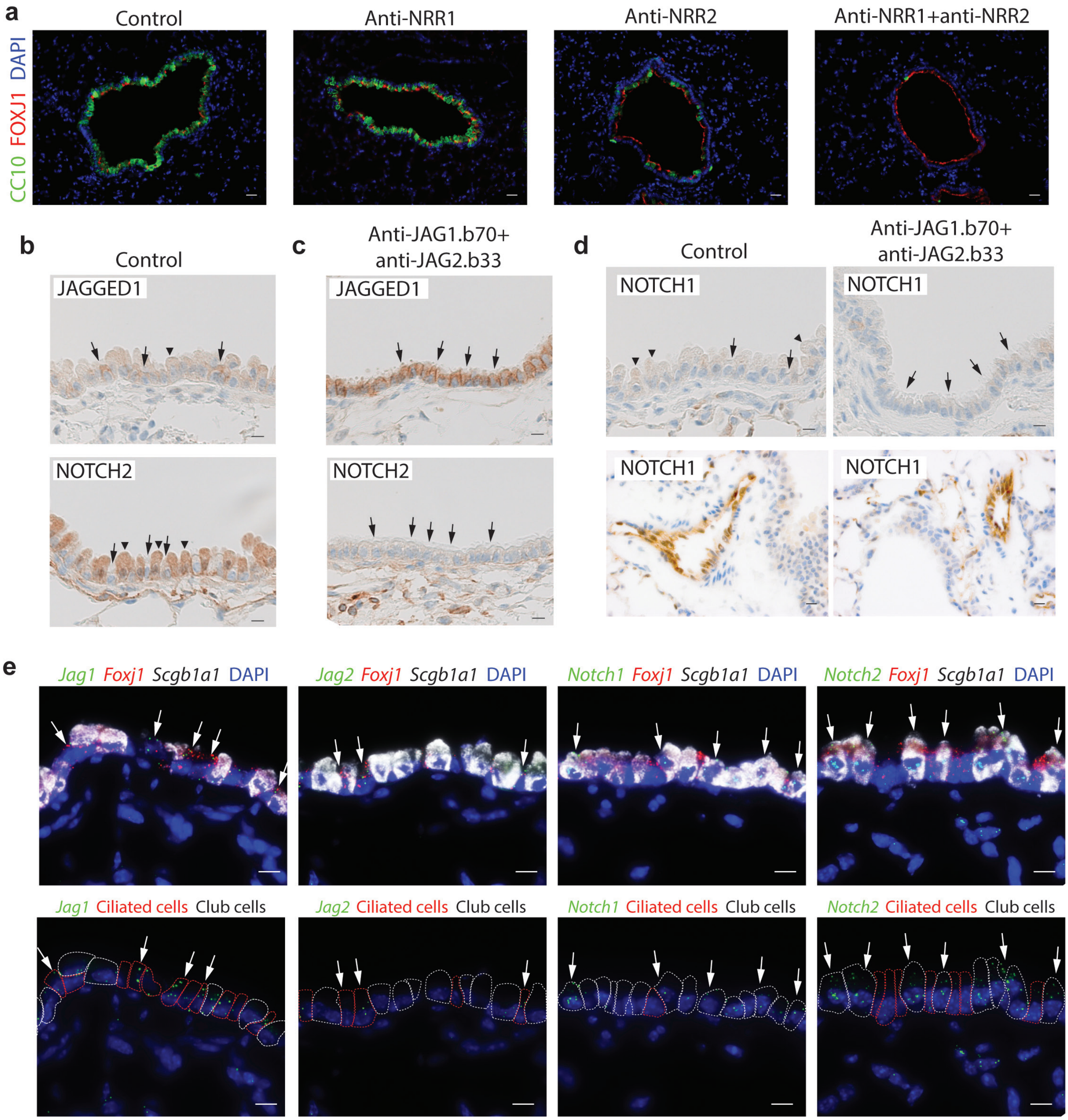
Extended Data Figure 2 | Notch ligand proteins used to characterize blocking antibodies, crystallographic data, and *in vitro* verification of lack of significant cross-reactivity of anti-JAG2.b33. a, Protein reagents used to characterize anti-JAG1 and anti-JAG2 binding. Purified human and mouse JAG1, JAG2, DLL1 and DLL4 extracellular fragments (2–5 μg) used for antibody characterization studies were analysed by SDS–PAGE under non-reducing (NR) and reducing (R) conditions. **b**, Table of crystallographic data collection and refinement statistics. **c**, Immunoblot analysis of NOTCH2 intracellular domain (NICD2) in U87 glioblastoma cells to assess selectivity of JAG1 inhibition. JAG1 immobilized on beads was used to induce NOTCH2 signalling in U87 cells, which endogenously express high levels of NOTCH2, in the presence of the indicated reagents. NOTCH2 signalling was assessed using an antibody that preferentially recognizes the γ-secretase-cleaved (active) form of NICD2 (ref. 33). As a control, a γ-secretase inhibitor (DAPT, 5 μM in DMSO) inhibited signalling relative to control treatment (DMSO alone), evidenced by a clear decrease in NICD2 levels. Likewise, anti-JAG1.b70 (25 μg ml^{-1}) completely blocked NOTCH2 signalling; by contrast, a high concentration of anti-JAG2.b33 (25 μg ml^{-1}) did not detectably decrease NICD2 levels, consistent with our other data that this antibody does not inhibit JAG1 signalling despite low-affinity binding to JAG1. Asterisk indicates a nonspecific band. A full scan of the blot may be found in Supplementary Fig. 1.



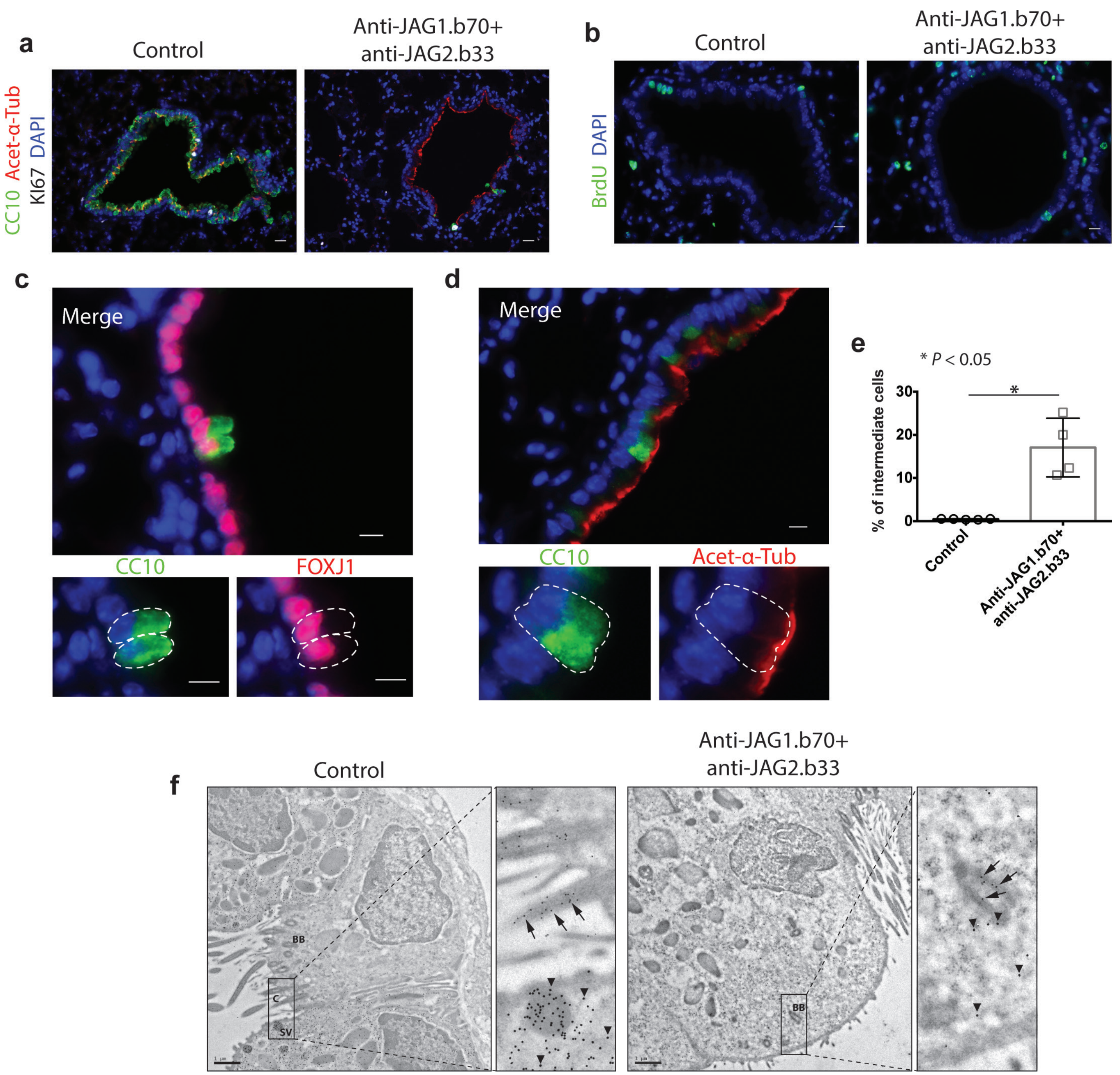
Extended Data Figure 3 | Anti-JAG1.b70 human–mouse cross-reactivity and inhibitory mechanism. a, **b**, Amino acid sequence alignment of human and mouse JAG1 (**a**) and human JAG1 and JAG2 (**b**) showing the DSL through EGF3 domains. Asterisk (*) indicates identical amino acids; colon (:) denotes conservative difference; full stop (.) denotes semi-conservative difference; and blank space () denotes non-conservative difference. **c**, The anti-JAG1.b70 epitope on human JAG1 is highlighted pink, with residues unique to mouse JAG1 highlighted in green. Similarly, residues differing between JAG1 and JAG2 positions are highlighted yellow. **d**, The NOTCH1–DLL4 crystal structure (PDB 4XL1; ref. 16) was superimposed onto our JAG–anti-JAG1-Fab coordinates. The high structural similarity between DLL4 and JAG1 is evident within the DSL and EGF1 domains (root mean squared deviation (r.m.s.d.) < 1.1 Å). In this view, the DLL4 C2 domain and the anti-JAG1 Fab were omitted for clarity. **e**, A ~90° view relative to **d**, with the anti-JAG1 Fab shown in space-filling representation. Depicting the anti-JAG1.b70 Fab bound to JAG1 indicates that the antibody light chain (LC; cyan) would clash with Notch1 (arrow), thus supporting an inhibitory mechanism based on steric occlusion of receptor binding.



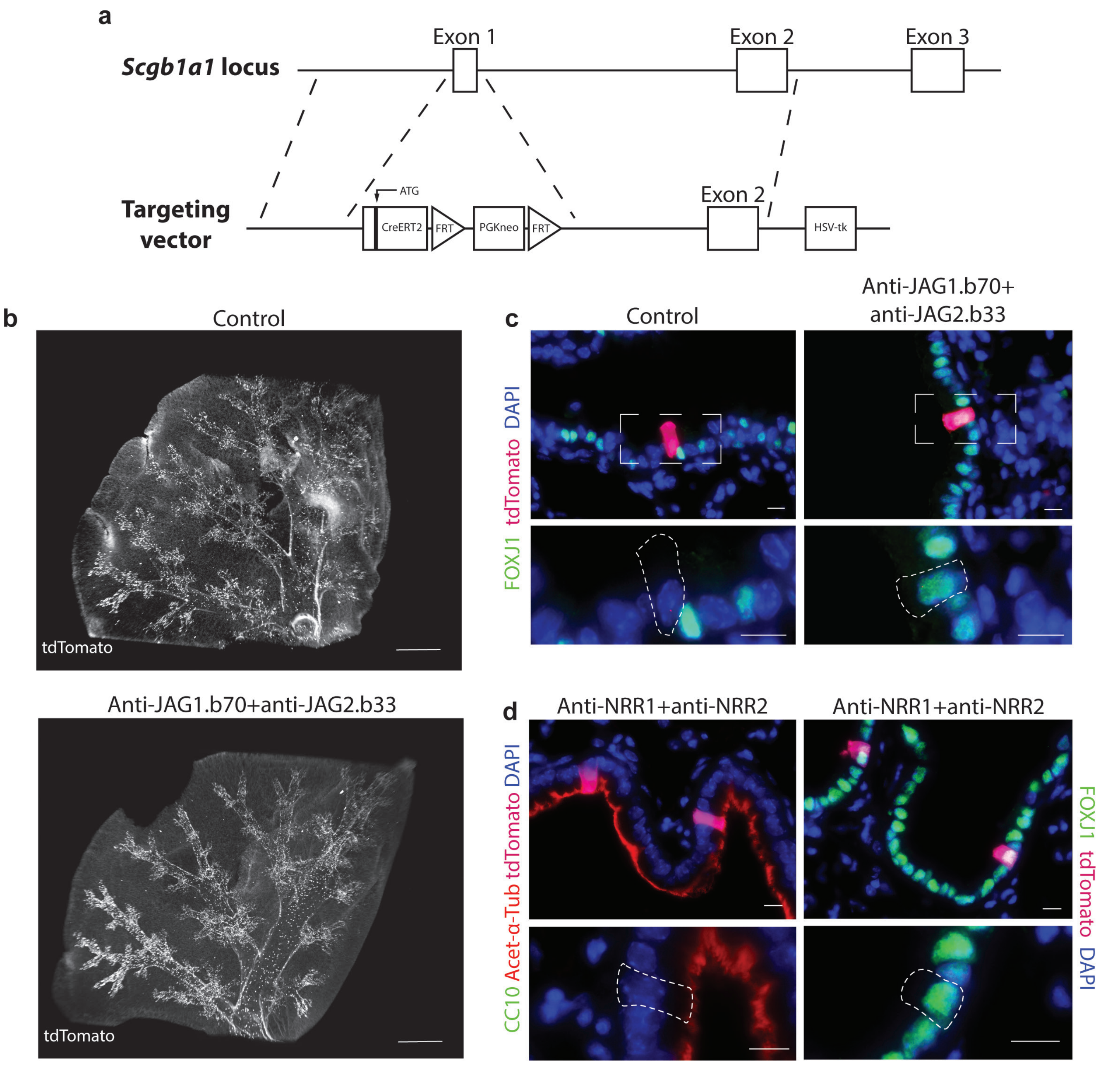
Extended Data Figure 4 | JAG blocking antibodies inhibit JAG-induced Notch signalling *in vivo*. a, Immunofluorescence staining of bronchiolar epithelium for CC10 (green) and FOXJ1 (red) from mice dosed every 3 days over 8 days with the indicated antibodies. Nuclei were counterstained with DAPI. **b**, Haematoxylin and eosin staining of skin from mice treated with control antibody, anti-JAG1.b70, anti-JAG2.b33 or anti-JAG1.b70 plus anti-JAG2.b33. JAG2 but not JAG1 inhibition induced a loss of mature sebocytes in sebaceous glands (arrows in control and anti-JAG1.b70 skin, left two panels) ($n = 3$ for each group). **c**, Immunofluorescence staining of bronchiolar epithelium for CC10 (green) and acetylated-α-tubulin (red) from mice treated with intranasal administration of control or anti-JAG1.b70 plus anti-JAG2.b33 antibodies for 5 days. Nuclei were labelled with DAPI ($n = 3$ for each group). **d**, Immunofluorescence staining for the acetylated-α-tubulin (red), the neuroepithelial-cell-specific marker CGRP (pink) and CC10 (green) from mice treated with anti-JAG1.b70 plus anti-JAG2.b33. Right panel shows a merged image and includes DAPI labelling. Club cells adjacent to neuroepithelial cells escape the JAG-blockade-induced transdifferentiation. **e**, Expression analysis of whole lungs from mice treated with the indicated antibodies ($n = 3$ for each group). Genes analysed included the Notch signalling targets *Hes1*, *Hey1* and *Hey2* as well as the club-cell-specific genes *Scgb1a1*, *Muc5b*, *Muc5ac* and the ciliated-cell-specific gene *Foxj1*. Both fold change and P values are plotted. Vertical and red horizontal dashed lines delineate twofold changes and $P < 0.05$, respectively. Anti-JAG1.b70 significantly increased expression of *Foxj1* and reduced expression of *Scgb1a1* by at least twofold. By contrast, no significant changes were observed when treating with anti-JAG2.b33 alone. The combination of anti-JAG1.b70 plus anti-JAG2.b33 significantly reduced expression of all Notch target genes analysed as well as club-cell-specific genes while increasing expression of *Foxj1*. Thus, the magnitude and type of gene expression changes revealed inhibition of Notch signalling and cell fate conversion in a manner that mirrored antibody induced cell fate changes assessed by immunofluorescence and other methods. Scale bars, 10 μm (**a**, **c**), 100 μm (**b**),



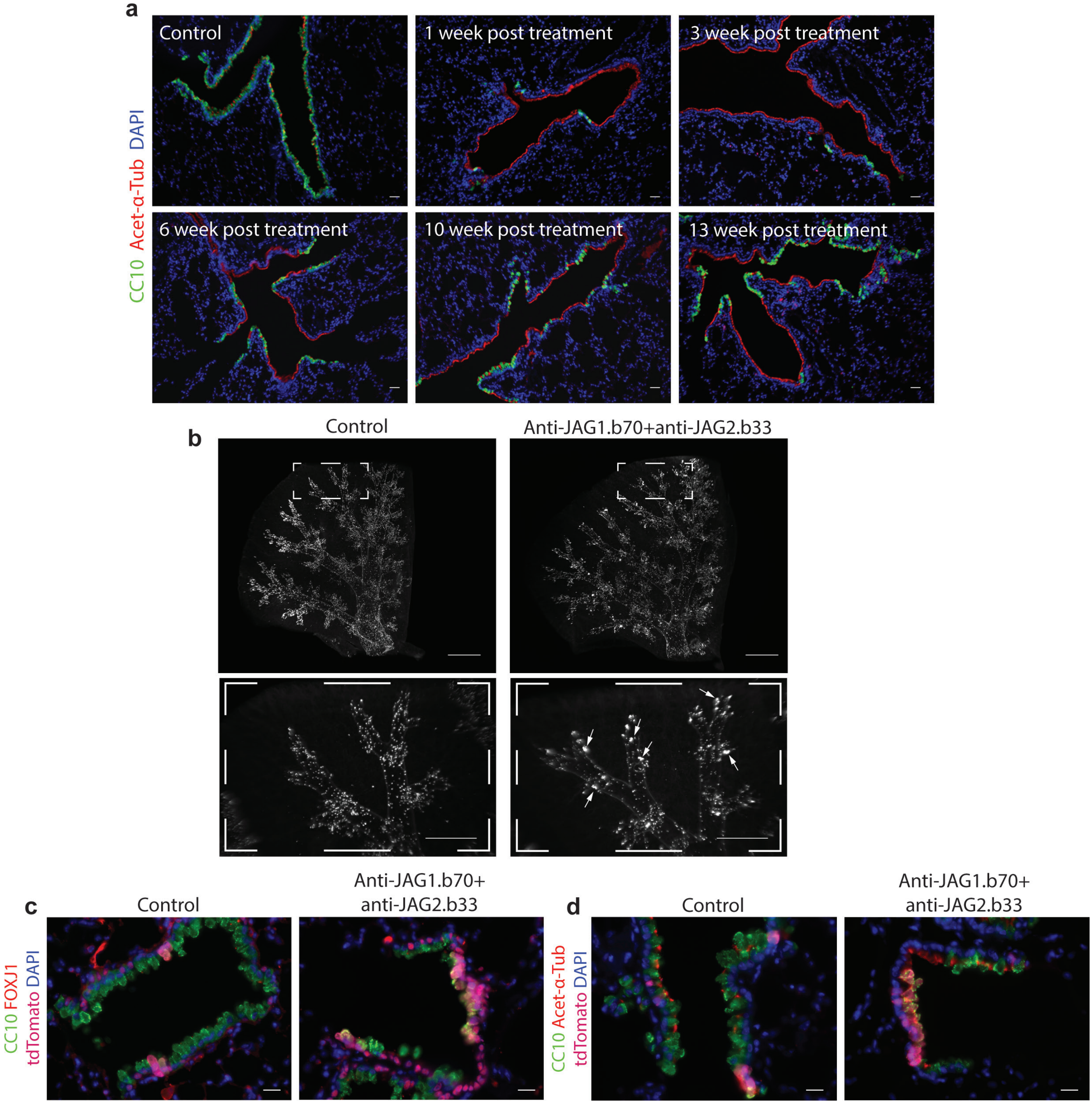
Extended Data Figure 5 | Determination of the functionally relevant ligand-receptor pair in the adult airway. a, Immunofluorescence staining for CC10 (green) and acetylated-α-tubulin (red) of mice treated with isotype control antibody, the NOTCH1 or NOTCH2 blocking antibodies anti-NRR1 or anti-NRR2, or the combination of anti-NRR1 plus anti-NRR2 ($n = 3$ for each group). NOTCH2 inhibition alone markedly reduced club cells numbers, with a corresponding increase in ciliated cell numbers, whereas NOTCH1 inhibition alone had little or no effect; dual blockade of both receptors induced the strongest phenotype. Results indicate that NOTCH2 is the dominant receptor controlling this phenotype, with NOTCH1 perhaps playing a secondary and functionally redundant role. **b**, **c**, Immunohistochemical staining of airways for JAG1 and NOTCH2 after control treatment (**b**) or JAG blockade (**c**) as in **a** revealed that JAG1 protein was most prominently expressed in ciliated cells, although a weak signal was observed throughout the epithelium. The JAG1 signal was strongly detected throughout the ciliated epithelium after JAG blockade, suggesting that the ciliated cells are the relevant signal-sending cell type. NOTCH2 protein staining showed a complementary pattern, with expression clearly localized to club cells and little or no signal detectable in the ciliated epithelium after JAG inhibition. **d**, Consistent with its secondary role relative to NOTCH2, NOTCH1 protein expression was very weak in the epithelial layer (top), with strong NOTCH1 staining on blood vessels from the same tissue sections serving as a positive control (bottom). **e**, Fluorescent RNAscope *in situ* hybridization to detect mRNA expression of *Jag1*, *Jag2*, *Notch1* and *Notch2* in mouse bronchiolar epithelium. Specific detection of each of the mouse NOTCH receptor and ligand mRNAs is shown in green, whereas mouse *Foxj1* and *Scgb1a1* mRNA, to mark ciliated and club cells, are shown in red and white, respectively. The signals using this method appear as coloured puncta. Samples were counterstained with DAPI to reveal nuclei. Consistent with their functioning as the primary ligand–receptor pair controlling cell fate, *Jag1* and *Notch2* were more highly expressed than *Notch1* and *Jag2*, which was only weakly detectable ($n = 4$ for each probe). *Jag1* and *Jag2* signals appear in the same cells as *Foxj1*, indicating co-expression in ciliated cells (arrows, two left panels). By contrast, *Notch1* and *Notch2* signals appear in the same cells as *Scgb1a1*, indicating co-expression in club cells (arrows, two right panels). These results thus confirm the immunohistochemistry findings and extend the expression results to include *Jag2*, for which a reliable immunohistochemistry method is lacking. Scale bars, 20 μm (**a**), 10 μm (**b**–**e**).



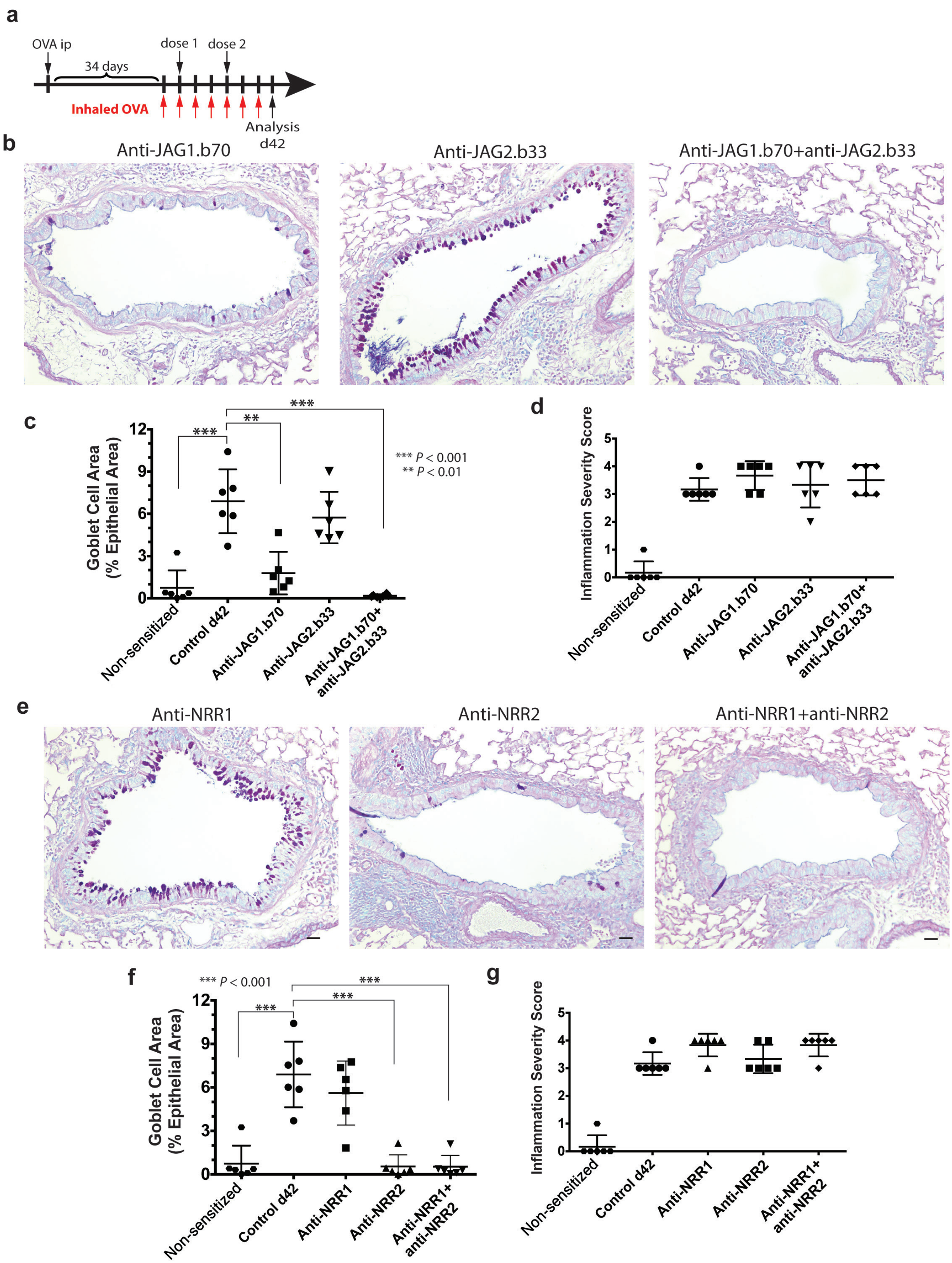
Extended Data Figure 6 | Lack of proliferation and intermediate cells are seen during transdifferentiation. a, Representative immunofluorescence images of sections used for quantification shown in Fig. 2f, of the percentage of airway cells stained for the proliferation marker KI67 (white) ($n = 5$ mice per group). **b**, Representative immunofluorescence images (BrdU, green) of sections used for the quantification shown in Fig. 2h, of percentages of BrdU-positive airway cells ($n = 5$ mice per group). **c**, **d**, Immunofluorescence staining for CC10 (green), and acetylated-α-tubulin (red, **c**) or FOXJ1 (red, **d**) of bronchiolar epithelium from mice 5 days after treatment with anti-JAG1.b70 plus anti-JAG2.b33. **e**, Quantification of the percentage of intermediate cells over the number of CC10-positive cells, that appeared positive for both CC10 and FOXJ1 by immunofluorescence at day 5 of control treatment or JAG blockade ($n = 4$, mean $\pm$ s.d.; unpaired t-test, $*P < 0.05$). A significant percentage of club cells remaining after JAG blockade ($17.0 \pm 6.8\%$) expressed both CC10 and FOXJ1, a master transcription factor dictating ciliated differentiation. Such CC10$^+$/FOXJ1$^+$ cells were also detected in control lungs, but at significantly lower percentages ($0.46 \pm 0.09\%$). **f**, Immunogold transmitted electron microscopy for high-resolution detection of CC10 (18-nm gold particles, arrowheads) and acetylated-α-tubulin (12-nm gold particles, arrows) in cells from the bronchiolar epithelium. Images of control cells (left) consistently showed CC10 expression restricted to secretory vesicles of apparent club cells, adjacent to and distinct from ciliated cells that expressed acetylated-α-tubulin in the basal bodies and cilia, confirming the specificity of immunostaining. Treatment with anti-JAG1.b70 plus anti-JAG2.b33 yielded a fraction of cells that displayed both acetylated-α-tubulin at basal bodies and CC10 in the cytoplasm (right), consistent with a phenotypic intermediate expected during a club-to-ciliated cell conversion. Scale bars, 20 μm (**a**, **b**), 10 μm (**c**, **d**) and 1 μm (**f**).



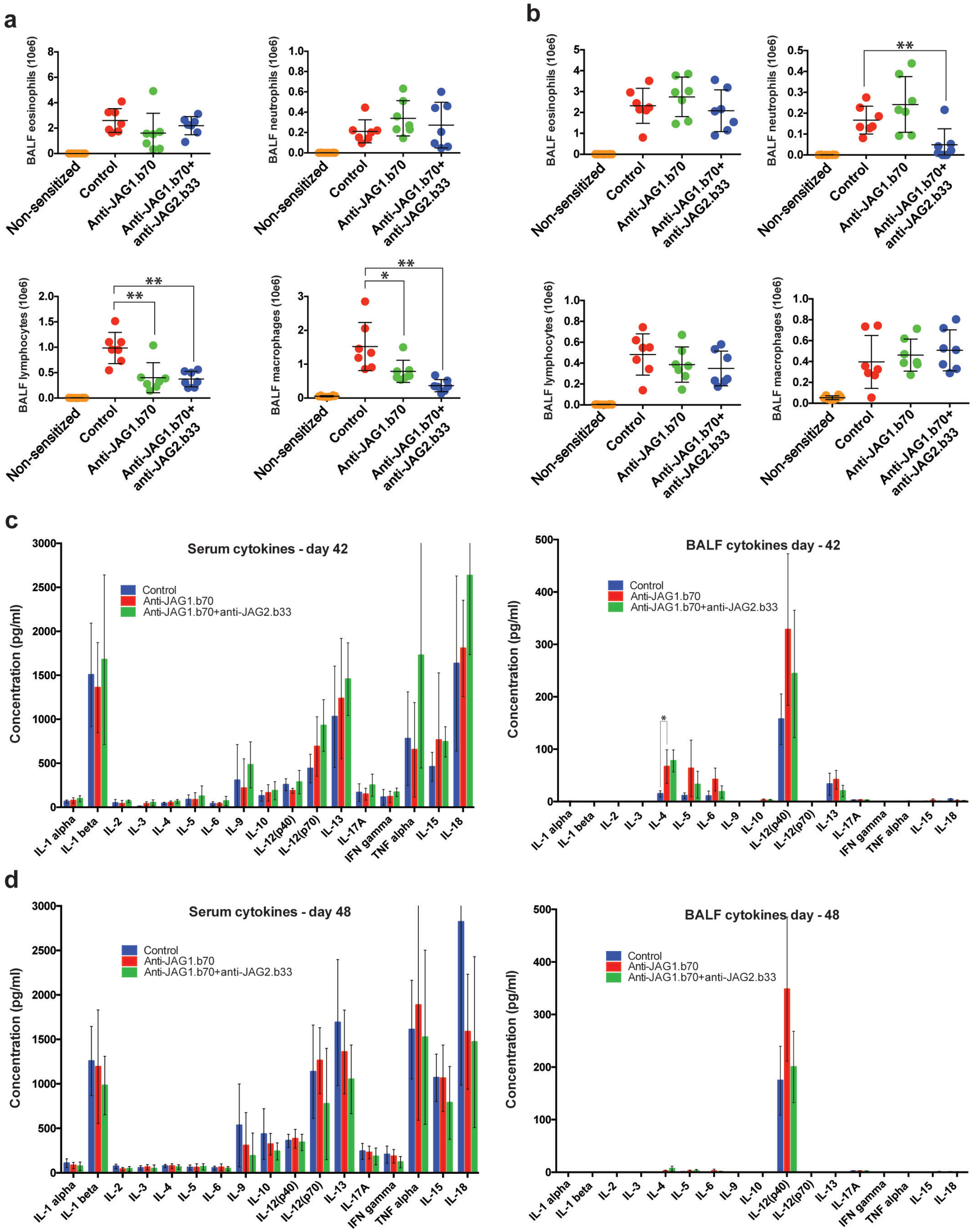
Extended Data Figure 7 | Lineage tracing shows that club cells transdifferentiate into ciliated cells. a, Targeting construct to generate Scgb1a1-CreERT2^{GNE} knock-in mice. PGK-neo and HSV-tk cassettes were used for positive and negative selection, respectively. See Methods for details. **b**, Whole-mount imaging of the right caudal lobe of Scgb1a1-CreERT2^{GNE}/Rosa26-lsl-tdTomato mice after mice were treated as in Fig. 3a. Images shown are maximum *z*-projections of optical sections obtained with an ultramicroscope ($n = 3$ for each group). Scale bar, 1 mm. The overall signal from the club cell lineage trace appears approximately equal and well distributed throughout the lobe, even after conversion to ciliated cells (bottom), confirming the lack of any notable cell loss, including apoptosis, in transdifferentiating club cells. **c**, Lineage tracing as in Fig. 3b, except using FOXJ1 instead of acetylated-α-tubulin as the ciliated cell marker. Scale bar, 10 μm. **d**, Lineage tracing results demonstrating that inhibition of NOTCH1 plus NOTCH2 induces club-to-ciliated cell transdifferentiation. Scale bar, 10 μm. After treating mice with anti-NRR1 and anti-NRR2 blocking antibodies, immunofluorescence staining of the bronchiolar epithelium was performed as in **c** and Fig. 3b. After NOTCH1 plus NOTCH2 inhibition, the tdTomato-positive cells, marking the club cell lineage, express acetylated-α-tubulin (left) and FOXJ1 (right), and have thus assumed a ciliated cell identity.



Extended Data Figure 8 | Club cells slowly reappear after antibody washout with expansion localized at the brochoalveolar duct junctions. **a**, Re-establishment of normal club and ciliated cell patterning after transdifferentiation. Immunofluorescence staining for CC10 (green) and acetylated-α-tubulin (red) of mice treated with a single dose of isotype control antibody or anti-JAG1.b70 plus anti-JAG2.b33. Bronchiolar epithelia were analysed 1, 3, 6, 10 and 13 weeks after treatment, as indicated, to determine the time needed for club cells to reappear ($n = 3$ for each time point). The first signs of recovery were evident after 6 weeks, with increased but incomplete recovery observed after 13 weeks. Reappearing rows of club cells seemed to originate from brochoalveolar duct junctions (BAJs), resulting in a gradient of reestablishment from smaller to larger airways. **b**, Whole-mount imaging of the right caudal lobe of Scgb1a1-CreERT2^{GNE}/Rosa26-lsl-tdTomato mice. After induction of lineage tracing with four doses of tamoxifen (200 mg kg^{-1}), mice were treated with a single dose of isotype control antibody (left) or anti-JAG1.b70 plus anti-JAG2.b33 (right), 1 week after the last tamoxifen injection. Lungs were analysed 13 weeks after treatment. Images shown are maximum *z*-projections of optical sections obtained with an Ultramicroscope ($n = 4$ for each group). In lungs from mice treated with anti-JAG1.b70 plus anti-JAG2.b33, clusters of lineage-traced cells were observed at the BAJs (arrows in bottom right panel); such clusters are absent in control lungs. **c**, **d**, Immunofluorescence staining of bronchiolar epithelium for the lineage-tracing marker tdTomato (pink), CC10 (green) and FOXJ1 (red, **c**) or acetylated-α-tubulin (red, **d**), from same mice as in **b**. Whereas mostly single cells and small clones of lineage-traced cells are found at the BAJs of control lungs, large clones are seen in lungs after JAG blockade, confirming the pattern of reestablishment of club cells after treatment (**a**). Nuclei were labelled with DAPI. Scale bars, 20 μm (**a**), 1 mm (**b**, top), 0.5 mm (**b**, bottom) and 10 μm (**c**, **d**).



Extended Data Figure 9 | NOTCH blockade inhibits goblet cell metaplasia *in vivo*. **a**, Mice were sensitized during a 35-day period after an i.p. injection of ovalbumin or vehicle (non-sensitized control) and then challenged with aerosolized ovalbumin for 7 consecutive days to induce inflammation and goblet cell metaplasia. Mice were also treated, at 1 and 4 days after challenge, with isotype control antibody, anti-JAG1.b70, anti-JAG2.b33 or anti-JAG1.b70 plus anti-JAG2.b33, as indicated ($n = 6$ mice per group, mean ± s.d.). **b**, Alcian blue/PAS staining of lung sections from mice treated with anti-JAG1.b70, anti-JAG2.b33 or the combination, on days 36 and 39, indicated as dose 1 and 2 in **a**. **c**, Quantification of goblet cell area ($n = 6$ for each group, mean ± s.d.; unpaired t-test, $***P < 0.001$, $**P < 0.01$). **d**, Inflammation index as assessed by haematoxylin and eosin staining of lung sections. **e**, Alcian blue/PAS staining of lung sections from mice treated with anti-NRR1, anti-NRR2, or the combination, on days 36 and 39, indicated as dose 1 and 2 in **a**. **f**, Quantification of goblet cell area ($n = 6$ for each group, mean ± s.d.; unpaired t-test, $***P < 0.001$). **g**, Inflammation index as assessed by haematoxylin and eosin staining of lung sections. All scale bars, 20 μm.



Extended Data Figure 10 | Characterization of the immune response during the ovalbumin challenge. a, Total numbers of immune cell populations found in the bronchoalveolar lavage fluid (BALF) of mice from the prevention goblet cell metaplasia study (Extended Data Fig. 9a–d). A significant reduction in both lymphocytes and macrophages is observed ($n = 7$ for each group, mean ± s.d.; unpaired t-test, $*P < 0.05$, $**P < 0.01$), but not in neutrophils and eosinophils, which are the most relevant cells that drive the metaplasia phenotype. **b**, Total numbers of immune cell populations found in the BALF of mice from the intervention goblet cell metaplasia study in Fig. 4. Although dual JAG blockade as well as blockade of JAG1 alone reverse goblet cell metaplasia (Fig. 4), a significant reduction in neutrophils was observed only after dual JAG blockade; no changes in other cell types, including eosinophils, were observed ($n = 7$ for each group, mean ± s.d.; unpaired t-test, $**P < 0.01$). **c**, Analysis of cytokine levels in blood serum (left) and BALF (right) of mice from the prevention study summarized in Extended Data Fig. 9a–d. Both antibody treatments resulted in a significant increase in the levels of IL4 in the BALF, although this increase was modest and not sustained to the later time point ($n = 7$ for each group; mean ± s.d.; unpaired t-test, $*P < 0.05$). **d**, Analysis of cytokine levels in blood serum (left) and BALF (right) of mice from the intervention study summarized in Fig. 4. Neither antibody treatment resulted in altered cytokine levels at this time point ($n = 7$ for each group; mean ± s.d., unpaired t-test). These results support an epithelial-cell-specific mechanism in which prevention and reversal of goblet cell metaplasia reflect direct effects of JAG inhibition on lung cell fate.

LETTER

doi:10.1038/nature16141

A mechanism for expansion of regulatory T-cell repertoire and its role in self-tolerance

Yongqiang Feng[1]*, Joris van der Veeken[1]*, Mikhail Shugay[2,3,4], Ekaterina V. Putintseva[2,4], Hatice U. Osmanbeyoglu[5], Stanislav Dikiy[1], Beatrice E. Hoyos[1], Bruno Moltedo[1], Saskia Hemmers[1], Piper Treuting[6], Christina S. Leslie[5], Dmitriy M. Chudakov[2,3,4] & Alexander Y. Rudensky[1]

T-cell receptor (TCR) signalling has a key role in determining T-cell fate. Precursor cells expressing TCRs within a certain low-affinity range for complexes of self-peptide and major histocompatibility complex (MHC) undergo positive selection and differentiate into naive T cells expressing a highly diverse self-MHC-restricted TCR repertoire. In contrast, precursors displaying TCRs with a high affinity for 'self' are either eliminated through TCR-agonist-induced apoptosis (negative selection)[1] or restrained by regulatory T (T_{reg}) cells, whose differentiation and function are controlled by the X-chromosome-encoded transcription factor Foxp3 (reviewed in ref. 2). Foxp3 is expressed in a fraction of self-reactive T cells that escape negative selection in response to agonist-driven TCR signals combined with interleukin 2 (IL-2) receptor signalling. In addition to T_{reg} cells, TCR-agonist-driven selection results in the generation of several other specialized T-cell lineages such as natural killer T cells and innate mucosal-associated invariant T cells[3]. Although the latter exhibit a restricted TCR repertoire, T_{reg} cells display a highly diverse collection of TCRs[4–6]. Here we explore in mice whether a specialized mechanism enables agonist-driven selection of T_{reg} cells with a diverse TCR repertoire, and the importance this holds for self-tolerance. We show that the intronic *Foxp3* enhancer conserved noncoding sequence 3 (CNS3) acts as an epigenetic switch that confers a poised state to the *Foxp3* promoter in precursor cells to make T_{reg} cell lineage commitment responsive to a broad range of TCR stimuli, particularly to suboptimal ones. CNS3-dependent expansion of the TCR repertoire enables T_{reg} cells to control self-reactive T cells effectively, especially when thymic negative selection is genetically impaired. Our findings highlight the complementary roles of these two main mechanisms of self-tolerance.

TCR signalling plays an essential role in T_{reg} cell differentiation and function[7–10]. Previous studies have shown that a broad range of self-reactivity can promote T_{reg} cell differentiation in the thymus, consistent with the highly diverse TCR repertoire of these cells[4–6]. We reasoned that a dedicated mechanism, linked to the regulation of *Foxp3* gene expression, might enable selection of T_{reg} cells with a diverse TCR repertoire. Previously, we showed that an intronic element of the *Foxp3* gene, CNS3, increases the efficiency of T_{reg} cell generation, raising the possibility that it might affect the composition of the T_{reg} TCR repertoire. To account for the potential effects of a mixed 129/B6 genetic background in our previous study, we backcrossed the CNS3 knockout $Foxp3^{\Delta CNS3\text{-}gfp}$ allele onto a B6 genetic background and generated male $Foxp3^{\Delta CNS3\text{-}gfp}$ and $Foxp3^{gfp}$ littermates carrying identical amino-terminal enhanced green fluorescent protein (GFP) reporters[11,12]. Consistent with our previous observation[11], we found an ~40% reduction in $Foxp3^{+}CD4^{+}$ thymocytes in CNS3-deficient mice, compared to CNS3-sufficient littermate controls ($2.05 \pm 0.38\%$ and $3.38 \pm 0.70\%$ (mean ± s.d.) of CD4 single-positive (SP) thymocytes, respectively). The size of other thymocyte subsets was unaffected (Extended Data Fig. 1a, b). In contrast, peripheral T_{reg} cells were present at comparable frequencies, probably owing to homeostatic expansion[7,13–15] (Extended Data Fig. 1a). Interestingly, loss of CNS3 had no effect on Foxp3 expression in differentiated T_{reg} cells (Extended Data Fig. 1c). Our previous study suggested that CNS3 is epigenetically marked in precursor cells, raising the question of which stage of T-cell differentiation CNS3 acts to facilitate T_{reg} cell development. We found that ablation of a conditional CNS3 allele in double positive (DP) or double negative (DN) thymocytes using $Cd4^{Cre}$ or Lck^{Cre} drivers, respectively, resulted in similarly defective thymic T_{reg} cell generation (Extended Data Fig. 1d, e). To assess the requirement for CNS3 immediately preceding Foxp3 induction, we acutely ablated CNS3 using tamoxifen-inducible Cre and observed decreased Foxp3 induction upon activation of naive $CD4^{+}$ T cells in the presence of TGFβ and IL-2 (Extended Data Fig. 1f). Notably, in mature T_{reg} cells, CNS3 was fully dispensable for the maintenance of Foxp3 expression during cell division in the presence of pro-inflammatory cytokines (Extended Data Fig. 1g, h), and for their suppressor function *in vivo* (Extended Data Fig. 2).

These findings raised the question of how, mechanistically, CNS3 could selectively facilitate the initiation but not the maintenance of Foxp3 expression. To address this problem, we identified the stage of thymocyte differentiation at which the CNS3 region first acquires the characteristic features of a poised enhancer. We previously found that CNS3 is marked by lysine 4 mono-methylation of histone H3 (H3K4me1) in DP thymocytes[11]. Unexpectedly, we found increased H3K4me1 levels at CNS3 at the DN1 stage and in haematopoietic stem cells, comparable to the levels observed in DP thymocytes, CD4 SP thymocytes and naive $CD4^{+}$ and $CD8^{+}$ T cells (Fig. 1a–c and unpublished data). In contrast, CNS3 chromatin was not enriched for H3K4me1 in embryonic stem cells, macrophages or dendritic cells (Fig. 1b, c). These results indicate that the poised state of CNS3 is established at a very early stage of haematopoiesis, but is lost in 'non-T-cell' lineages. As CNS3 appeared to be the earliest epigenetically modified region in the *Foxp3* locus, it might exert its function by facilitating chromatin remodelling at the *Foxp3* promoter.

While deposition of the 'active' histone modifications H3K4me3 and H3K27ac at the *Foxp3* promoter occurred exclusively in T_{reg} cells (Extended Data Fig. 3a, b), we found an enrichment of H3K4me1 in mature CD4 SP thymocytes and naive $CD4^{+}$ T cells (Fig. 1d). In the absence of CNS3, both mature CD4 SP thymocytes and naive $CD4^{+}$ T cells showed impaired H3K4me1 accumulation at the *Foxp3* promoter (Fig. 1e, f), suggesting that CNS3 facilitates epigenetic remodelling of the *Foxp3* promoter in T_{reg} cell precursors. Notably, differentiated CNS3-deficient T_{reg} cells showed normal levels of H3K4me3 and

[1]Howard Hughes Medical Institute and Immunology Program, Ludwig Center at Memorial Sloan Kettering Cancer Center, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [2]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS, Miklukho-Maklaya 16/10, Moscow 117997, Russia. [3]Pirogov Russian National Research Medical University, Ostrovityanova 1, Moscow 117997, Russia. [4]Central European Institute of Technology, Masaryk University, Kamenice 753/5, Brno 62500, Czech Republic. [5]Computational Biology Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [6]Department of Comparative Medicine, School of Medicine, University of Washington, Seattle, Washington 98195, USA.
*These authors contributed equally to this work.

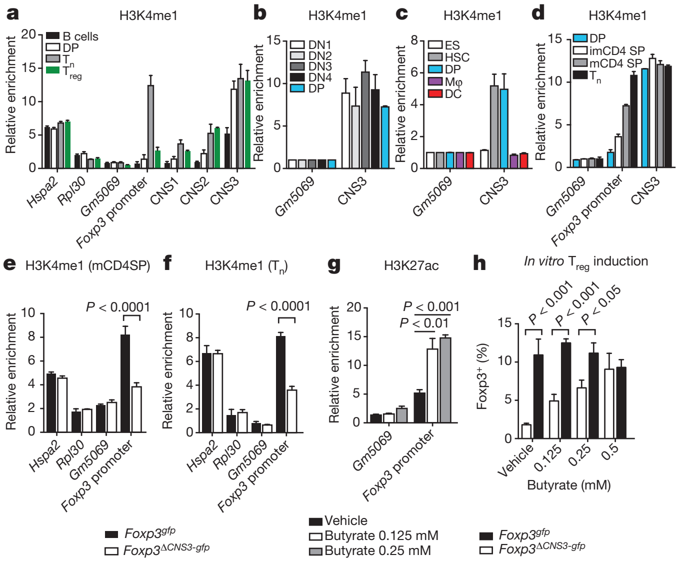


Figure 1 | CNS3 acts as an epigenetic switch for the *Foxp3* promoter poising. a, Chromatin immunoprecipitation and quantitative PCR (ChIP–qPCR) of H3K4me1 at the *Foxp3* locus and control loci (*Hspa2*, *Rpl30* and *Gm5069*) in B cells, DP thymocytes, naive $CD4^+$ T (T_n) and T_{reg} cells. **b**, **c**, H3K4me1 at CNS3 in DN and DP thymocytes (**b**), and haematopoietic stem cells (HSC), embryonic stem (ES) cells, macrophages (Mφ) and dendritic cells (DC) (**c**). **d**, H3K4me1 at the *Foxp3* promoter in DP, immature CD4 SP (imCD4SP; $Foxp3^-CD62L^{lo}CD69^{hi}$), mature CD4 SP (mCD4SP; $Foxp3^-CD62L^{hi}CD69^{lo}$) thymocytes, and naive $CD4^+$ T cells. **e**, **f**, CNS3 dependent H3K4me1 at the *Foxp3* promoter in mature CD4 SP thymocytes (**e**) and naive $CD4^+$ T cells (**f**). **g**, **h**, Histone deacetylase inhibitor butyrate enhances H3K27ac at the *Foxp3* promoter (**g**) and rescues impaired T_{reg} differentiation of CNS3-deficient T cells *in vitro* (**h**). Two-tailed unpaired *t*-test. Error bars, mean ± s.e.m.; data represent triplicate cultures in 1 of ≥2 experiments.

H3K27ac deposition at the *Foxp3* regulatory regions (Extended Data Fig. 3c–e), consistent with the dispensable role of CNS3 in differentiated T_{reg} cells (Extended Data Fig. 1c, g, h).

To address whether the CNS3-dependent poised state of the *Foxp3* promoter assists deposition of additional permissive marks and further chromatin remodelling that facilitates the initiation of Foxp3 expression, we cultured naive $CD4^+$ T cells from male *Foxp3*^{gfp} and *Foxp3*$^{\Delta CNS3\text{-}gfp}$ littermates under T_{reg} cell differentiation conditions, and isolated $Foxp3^-$ cells that had been exposed to Foxp3-inducing conditions but had not yet acquired Foxp3 expression. We observed a CNS3-dependent increase in H3K27ac at the *Foxp3* promoter preceding Foxp3 expression (Extended Data Fig. 3f), consistent with the defect in T_{reg} cell differentiation in the absence of CNS3 (Extended Data Fig. 1f). Furthermore, blocking the recruitment of bromodomain-containing histone acetylation readers using the inhibitor iBET sharply reduced T_{reg} cell induction efficiency in a dose-dependent manner[16] (Extended Data Fig. 3g). Conversely, blocking histone deacetylase activity using butyrate increased H3K27ac at the *Foxp3* promoter in agreement with recent reports[17–19] (Fig. 1g). Notably, provision of butyrate rescued impaired *in vitro* T_{reg} cell differentiation associated with loss of CNS3 (Fig. 1h). These observations suggest that a CNS3-dependent poised state at the *Foxp3* promoter, probably via looping, in precursor cells may enhance their sensitivity to Foxp3-inducing signals (Extended Data Fig. 3h).

While CNS3-dependent poising of the *Foxp3* promoter could facilitate Foxp3 induction in a probabilistic manner, it might also enable lower strength TCR signals to promote T_{reg} cell differentiation. To address this possibility, we tested whether impaired Foxp3 induction in CNS3-deficient naive $CD4^+$ T cells could be rescued by increasing amounts of CD3 antibody under *in vitro* T_{reg} cell differentiation conditions. We found that the relative difference in the efficiency of Foxp3 induction between CNS3-sufficient and -deficient $CD4^+$ T cells was markedly decreased in the presence of higher amounts of CD3 antibody (Fig. 2a and Extended Data Fig. 4a), suggesting that increased TCR signal strength can partially compensate for the lack of *Foxp3* promoter poising in the absence of CNS3, and that differentiation of T_{reg} precursors receiving lower TCR stimulation might be disproportionally impeded by CNS3 deficiency. These results suggest that the mature T_{reg} cells differentiated from CNS3-deficient precursors are enriched for TCRs at the higher end of the self-reactivity spectrum, and depleted of those with lower self-reactivity.

To test this possibility we analysed $Foxp3^-$ and $Foxp3^+$ CD4 SP thymocytes and naive $CD4^+$ T cells for the expression of orphan nuclear receptor Nur77 (also known as Nr4a1), the product of a prominent TCR target gene in T_{reg} and conventional T cells that accurately reports the strength of TCR signalling[20,21]. In agreement with the *in vitro* Foxp3 induction studies, we found that both thymic and peripheral CNS3-deficient T_{reg} cells, but not $Foxp3^-CD4^+$ T cells, were markedly enriched for cells expressing higher levels of Nur77 in comparison to their CNS3-sufficient counterparts. This trend was observed in non-competitive settings of male *Foxp3*^{gfp} and *Foxp3*$^{\Delta CNS3\text{-}gfp}$ littermates, as well as in competitive settings, on ablation of CNS3 in bone marrow chimaeras and heterozygous female *Foxp3*$^{gfp/+}$ and *Foxp3*$^{\Delta CNS3\text{-}gfp/+}$ littermates (Fig. 2b, Extended Data Fig. 4b–d and unpublished data). Notably, Nur77 levels were increased in both resting ($CD44^{lo}CD62L^{hi}$) and activated ($CD44^{hi}CD62L^{lo}$) CNS3-deficient T_{reg} cells (Extended Data Fig. 4e–g). In contrast, Nur77 expression in $Foxp3^-$ thymocytes and peripheral T cells was unaffected in the absence of CNS3 (Extended Data Fig. 4d). Consistently, CNS3-deficient T_{reg} cells expressed increased amounts of CTLA4, a major negative feedback regulator of TCR signalling, and cell proliferation marker Ki-67 (Extended Data Fig. 4h, i). Finally, we found that CNS3-deficient and -sufficient $Foxp3^+$ CD4 SP thymocytes and resting peripheral T_{reg} cells exhibited distinct gene expression profiles, in contrast to $Foxp3^-$ subsets (Extended Data Fig. 5a). Specifically, the expression of TCR-dependent genes and genes characteristic of activated T_{reg} cells was significantly increased in CNS3-deficient $Foxp3^+$ thymocytes and resting T_{reg} cells in comparison to their CNS3-sufficient counterparts (Fig. 2c, d and Extended Data Fig. 5b, d). In contrast, transcriptional profiles of CNS3-sufficient and -deficient activated T_{reg} cells and $Foxp3^-$ CD4 SP thymocytes were similar (Extended Data Fig. 5a, c, e and unpublished data). These results further support the notion that loss of CNS3 results in enrichment of thymic T_{reg} cells with heightened TCR signal strength.

To assess the auto-reactivity of CNS3-deficient versus -sufficient T_{reg} cells *in vivo*, we examined their capacity for MHC class II (MHC-II)-dependent homeostatic expansion under lymphopenic conditions, known to be proportional to TCR affinity for self[5]. CNS3-deficient T_{reg} cells expanded markedly compared to CNS3-sufficient counterparts after co-transfer with congenically labelled effector T cells into lymphopenic hosts (*Tcrb*$^{-/-}$ *Tcrd*$^{-/-}$), probably driven by the recognition of self-antigens presented by MHC-II molecules because antibody-mediated blockage of MHC-II prevented expansion of T_{reg} cells and erased the advantage of CNS3-deficient T_{reg} cells over their CNS3-sufficient counterparts (Fig. 2e, f and Extended Data Fig. 5f). Accordingly, the frequency of CNS3-deficient T_{reg} cells was noticeably increased in the periphery of *Foxp3*$^{\Delta CNS3\text{-}gfp/+}$ compared to *Foxp3*$^{gfp/+}$ heterozygous female mice (Extended Data Fig. 5g). Thus, T_{reg} cells developed from precursors lacking CNS3 resulted in a skewed TCR repertoire.

We next examined the TCR repertoires of T_{reg} cells and naive and activated $CD4^+$ T cells in *Foxp3*$^{\Delta CNS3\text{-}gfp}$ or *Foxp3*^{gfp} *Tcra*$^{-/+}$ mice expressing the DO11.10 TCRβ chain transgene, through which TCR diversity is limited to a single functional TCRα chain locus[5]. Barcoded TCRα libraries were generated using an optimized protocol, and high-throughput sequencing data were analysed using the MIGEC software package[22]. Cluster analysis using VDJtools[23] showed

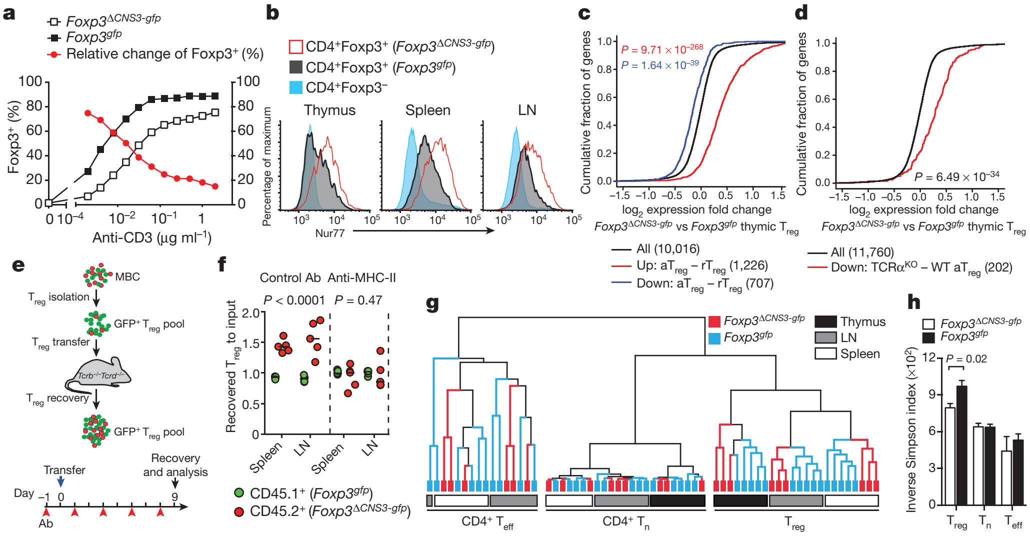


Figure 2 | CNS3 shapes the T_{reg} cell repertoire. a, CNS3 facilitates *in vitro* T_{reg} induction at suboptimal TCR signalling strength. Error bars, mean ± s.e.m. of triplicate cultures representing 1 of >2 experiments. **b**, Nur77 protein expression in CNS3-deficient and -sufficient T_{reg} cells (summarized in Extended Data Fig. 4b). LN, lymph nodes. **c**, **d**, Relative gene expression levels (RNA-seq) were compared to the upregulated (up) and downregulated (down) genes in activated (aT_{reg}) versus resting (rT_{reg}) T_{reg} cells from wild-type *Foxp3*^{gfp} mice (**c**), or to those downregulated in aT_{reg} cells subjected to Cre-induced TCR ablation or mock treatment (**d**). The numbers of genes in each comparison are indicated in parentheses. *Foxp3*^{gfp} ($n = 3$); *Foxp3*$^{\Delta CNS3\text{-}gfp}$ ($n = 4$). One-tailed Kolmogorov–Smirnov test. **e**, **f**, Analysis of the expansion potentials of CNS3-deficient and -sufficient T_{reg} cells in lymphopenic hosts. CD45.1^{+} and CD45.2^{+} T_{reg} cells sorted from mixed bone-marrow chimaeras (MBC) of CD45.1^{+} *Foxp3*^{gfp} and CD45.2^{+} *Foxp3*$^{\Delta CNS3\text{-}gfp}$ were mixed at a 1:1 ratio and co-transferred with wild-type naive Foxp3^{-}CD4^{+} T cells into *Tcrb*$^{-/-}$ *Tcrd*$^{-/-}$ recipients treated with MHC-II-blocking antibody (Ab) or control IgG before and after the transfer ($n = 5$ per group). Ratios of recovered T_{reg} cells to their inputs are shown (**f**). Two-tailed unpaired Mann–Whitney test, representative of 1 of 3 experiments. **g**, **h**, Cluster (**g**) and diversity (**h**) analysis of TCRα repertoires in *Foxp3*^{gfp} ($n = 5$) and *Foxp3*$^{\Delta CNS3\text{-}gfp}$ ($n = 3$) *Tcra*$^{-/+}$ mice bearing the DO11.10 TCRβ transgene. An identical sampling size was used to assess the diversity with the inverse Simpson index. Coloured bars represent individual mice. T_{eff}, effector CD4^{+} T cells. Error bars, mean ± s.e.m.; two-tailed unpaired *t*-test.

that TCRα repertoires of thymic and peripheral CNS3-deficient and -sufficient T_{reg} cells, but not naive or activated effector CD4^{+} T cells, were distinct (Fig. 2g). Further analysis showed a significantly reduced TCRα diversity of CNS3-deficient T_{reg} cells, but not naive or effector CD4^{+} T cells, in comparison to their CNS3-sufficient counterparts (Fig. 2h). As the TCR complementarity-determining region 3 (CDR3) largely determines TCR specificity for peptide–MHC (pMHC) complexes, we assessed the frequencies of strongly interacting amino acid residues in the TCRα chain CDR3 by leveraging a mathematical model linking the features of amino acid residues in the CDR3 to TCR affinity for pMHC[24]. Interestingly, the TCRα CDR3s were significantly enriched for strongly interacting amino acid residues (Extended Data Fig. 5h, i), and for more randomly added nucleotides (Supplementary Table 1) in CNS3-deficient versus -sufficient T_{reg} cells, but not naive or effector CD4^{+} T cells. These results implied higher affinities of CNS3-deficient T_{reg} TCRs for self-antigens and further supported the notion that CNS3 shapes T_{reg} TCR repertoire by increasing its diversity, probably by enabling T_{reg} differentiation in response to a broad range of self-reactivity.

To understand the functional significance of CNS3-dependent regulation of the T_{reg} cell repertoire, we first assessed the immune status of male *Foxp3*$^{\Delta CNS3\text{-}gfp}$ and their wild-type *Foxp3*^{gfp} littermates. CNS3 deficiency had no observed effect on the numbers of activated or memory CD4^{+} or CD8^{+} T cells, or on cytokine production by T cells in the secondary lymphoid organs of 8–12-week-old mice (Fig. 3a–c and unpublished data). Although CNS3-deficient T_{reg} cells were capable of preventing systemic autoimmunity (Fig. 3a–c and Extended Data Fig. 1a), it remained possible that the skewed T_{reg} TCR repertoire might have 'holes'. Thus, we reasoned that a select few non-lymphoid organs may exhibit focused immune activation in CNS3-deficient versus CNS3-sufficient mice, whereas others might be similarly or even more protected against autoimmunity by the over-represented highly autoreactive T_{reg} cells. Indeed, we found increased numbers of activated effector T cells and elevated IL-13 and IFNγ production by T cells in the lungs of *Foxp3*$^{\Delta CNS3\text{-}gfp}$ mice (Fig. 3a–c). We also observed markedly increased titres of circulating autoantibodies against several self-antigens in the sera of CNS3-deficient mice versus their wild-type *Foxp3*^{gfp} littermates, whereas the absolute amounts of immunoglobulin (Ig) isotypes were comparable (Fig. 3d, Extended Data Fig. 6a and unpublished data). This notion was further supported by the observed modest, but consistent, decrease in the severity of experimental autoimmune encephalomyelitis in *Foxp3*$^{\Delta CNS3\text{-}gfp}$ versus *Foxp3*^{gfp} littermates (Extended Data Fig. 6b–f). To compare the suppressive capacity of T_{reg} cells developed in CNS3-deficient and -sufficient mice, we transferred these Ly5.2^{+} T_{reg} cells together with Ly5.1^{+} *Foxp3*-null (Δ*Foxp3*) effector T cells into T-cell-deficient recipients (Fig. 3e). Despite comparable expansion of CNS3-sufficient and -deficient T_{reg} cells in the recipients, we observed more pronounced weight loss and increased pro-inflammatory cytokine production by effector Δ*Foxp3* T cells in the presence of CNS3-deficient T_{reg} cells in comparison to the control (Fig. 3f–h and Extended Data Fig. 6g, h). These results indicate that T_{reg} cells developed from CNS3-deficient precursors were selectively impaired in their capacity to suppress self-reactive effector T cells.

Both negative selection and T_{reg} cell generation are driven by self-antigen recognition in the thymus and probably have complementary roles in self-tolerance[1,2,25,26]. We reasoned that the relatively mild impairment in suppressive capacity of T_{reg} cells from CNS3-deficient mice on a B6 genetic background, resistant to autoimmunity, may not fully reveal the biological significance of CNS3-dependent broadening of the T_{reg} cell repertoire because of efficient negative selection. Therefore, we assessed the consequences of combined deficiency in CNS3 and *Aire* (autoimmune regulator), a nuclear factor required for thymic negative selection and optimal T_{reg} cell generation. Loss of *Aire* leads to diminished expression of a subset of tissue-restricted antigens in the thymus and, consequently, an enlarged self-reactive effector T-cell pool and diminished T_{reg} repertoire[25,27–29]. In contrast to late-onset and mild autoimmunity observed in *Aire*-knockout (KO)

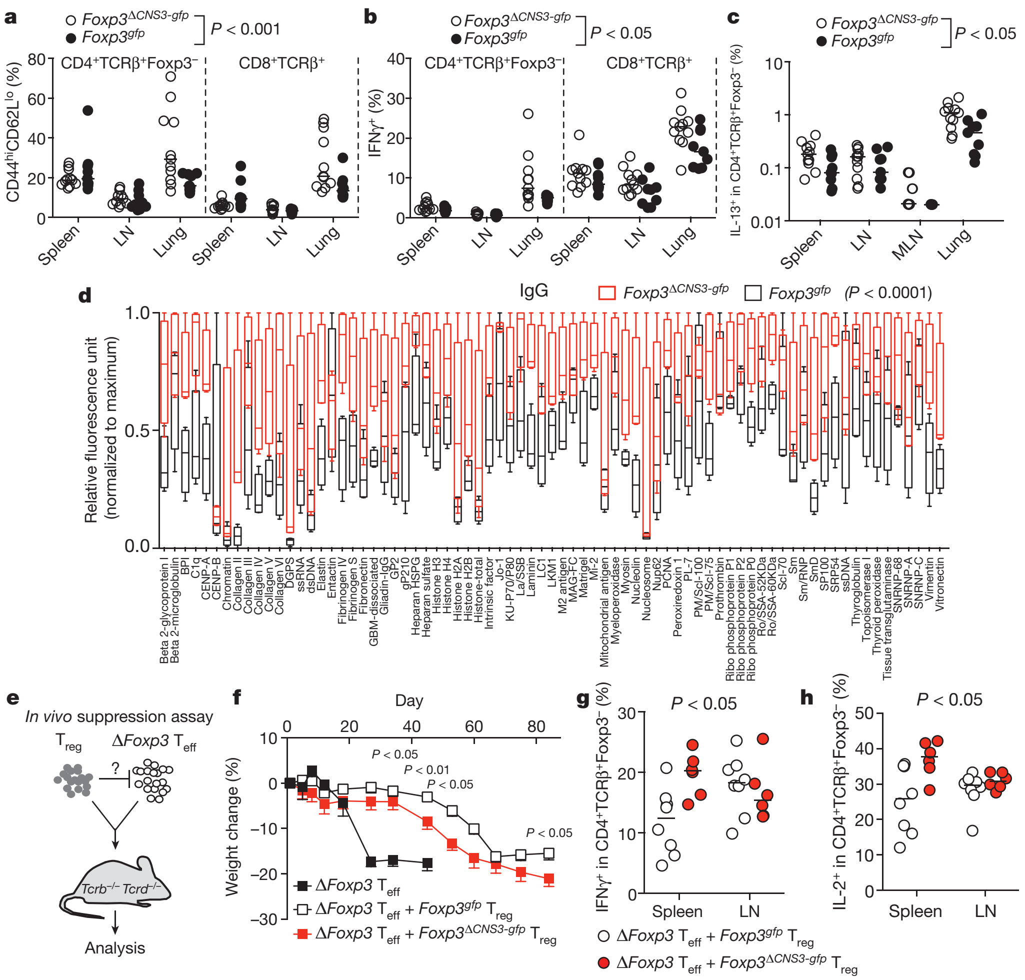


Figure 3 | Defective self-tolerance in the presence of CNS3-deficient T_{reg} cells. **a**–**c**, Analysis of the activation ($CD44^{hi}CD62L^{lo}$) (**a**), IFNγ (**b**) and IL-13 (**c**) production in $CD4^+Foxp3^-$ and/or $CD8^+$ T cells in *Foxp3*$^{\Delta CNS3\text{-}gfp}$ ($n = 11$) and *Foxp3*^{gfp} ($n = 9$) mice. Two-tailed unpaired Mann–Whitney test. MLN, mesenteric LN. **d**, Analysis of circulating IgG against multiple self-antigens in the serum of *Foxp3*^{gfp} and *Foxp3*$^{\Delta CNS3\text{-}gfp}$ littermates ($n = 4$ per group). Plots show minimum, maximum, first and third quartiles and median. Two-way analysis of variance (ANOVA). **e**–**h**, Compromised suppressor capacity of CNS3-deficient T_{reg} cells *in vivo*. $CD45.2^+$ CNS3-sufficient or -deficient T_{reg} cells were co-transferred at a 1:10 ratio with $CD45.1^+$ *Foxp3*-null effector $CD4^+$ T cells (Δ*Foxp3* T_{eff}) into *Tcrb*$^{-/-}$ *Tcrd*$^{-/-}$ mice (**e**). Recipient mice were monitored for body weight change (**f**), IFNγ (**g**) and IL-2 (**h**) production. Mice transferred with Δ*Foxp3* T_{eff} alone succumbed to severe inflammation and were killed on day 45. Δ*Foxp3* T_{eff} ($n = 5$); Δ*Foxp3* T_{eff} plus *Foxp3*^{gfp} ($n = 8$); Δ*Foxp3* T_{eff} plus *Foxp3*$^{\Delta CNS3\text{-}gfp}$ ($n = 6$). Two-tailed unpaired *t*-test (**f**) or Mann–Whitney test (**g**, **h**). Error bars, mean and s.e.m. (**f**). Results are representative of 2 independent experiments.

mice on a B6 genetic background, deficiency of both CNS3 and *Aire* resulted in fatal early-onset aggressive autoimmune lesions in multiple tissues as early as 3–4 weeks of age, whereas detectable autoimmune inflammation was lacking in littermates with a single deficiency in *Aire* or CNS3 (Fig. 4a and Extended Data Fig. 7a). We noticed a 100% ($n > 35$) penetrance with a stochastic gender-independent variation in manifestations expected from perturbations in randomly generated repertoires of self-reactive T cells as well as the probabilistic nature of negative selection[30] (Extended Data Fig. 7a and unpublished data). This was accompanied by significant increases in $CD4^+$ T-cell activation, IFNγ production (Fig. 4b, c), serum Ig levels (Extended Data Fig. 7b) and autoantibody production (Fig. 4d). Combined *Aire* and CNS3 deficiency resulted in a further reduction in thymic T_{reg} cell frequency in comparison to the single-deficient mice ($1.66 \pm 0.28\%$ and $0.91 \pm 0.35\%$ in *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/WT}$ and *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/KO}$ mice, respectively)[25] (Fig. 4e). However, peripheral T_{reg} cells reached normal levels in young *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/KO}$ mice before development of clinical signs of disease, probably owing to homeostatic proliferation (Fig. 4f). Despite their normal quantities and Foxp3 expression, these T_{reg} cells were unable to suppress pathogenic self-reactive T cells resulting from impaired negative selection in the absence of *Aire* (Fig. 4b–d and Extended Data Fig. 7c). As diminished thymic T_{reg} cell numbers and their skewed TCR repertoire probably contributed to disease severity in *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/KO}$ mice, we directly assessed the ability of CNS3-sufficient and -deficient T_{reg} cells developed in the presence of *Aire* to control *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/KO}$ effector T cells when adoptively transferred into T-cell-deficient hosts. Although the negative effect of CNS3 deficiency on the TCR repertoire was probably mitigated by T_{reg} cell expansion in lymphopenic settings, CNS3-deficient T_{reg} cells still exhibited compromised ability to suppress the responses of transferred *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/KO}$ effector T cells and resident B cells in comparison to the controls (Extended Data Fig. 7d–i). These results suggest that control of broad self-reactive T cells requires a diverse CNS3-dependent repertoire of T_{reg} cells.

Our studies suggest that CNS3, an intronic *Foxp3* regulatory element, establishes a poised state of the *Foxp3* promoter in precursor cells and increases the probability of Foxp3 induction in response to TCR stimulation, particularly within a lower range of signal strength (Extended Data Fig. 8). Similar mechanisms of promoter poising may operate in other cell types and enable them to respond to a wider spectrum of growth factor or morphogen concentrations through receptor-triggered analogue signalling. CNS3-mediated *Foxp3* promoter poising expands

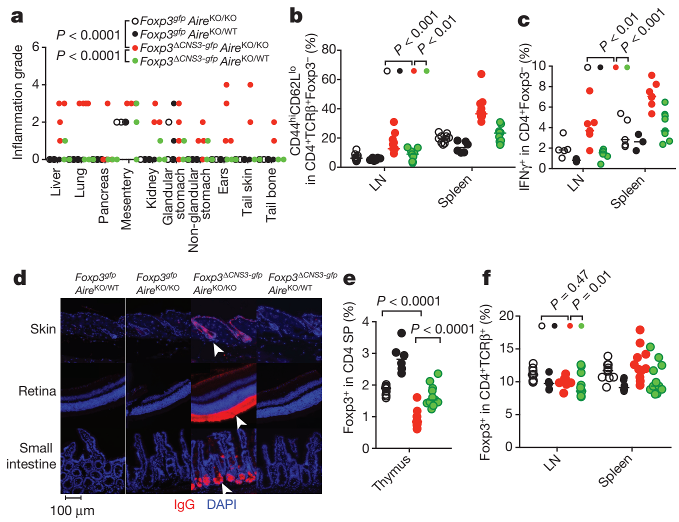


Figure 4 | CNS3-deficient T_{reg} cells fail to maintain self-tolerance in the absence of *Aire*. **a**, Analysis of tissue inflammation in CNS3 *Aire* double knockout (DKO) mice. $n = 3$ per group except for *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/KO}$ ($n = 4$). Two-tailed unpaired Mann–Whitney test. **b**, Analysis of activated $CD4^+$ T_{eff} cells in DKO mice. *Foxp3*^{gfp} *Aire*$^{KO/KO}$ ($n = 10$); *Foxp3*^{gfp} *Aire*$^{KO/WT}$ ($n = 6$); *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/KO}$ ($n = 11$); *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/WT}$ ($n = 11$). Two-tailed unpaired Mann–Whitney test. **c**, Analysis of IFNγ production by $CD4^+Foxp3^-$ T cells in DKO mice. *Foxp3*^{gfp} *Aire*$^{KO/KO}$ ($n = 5$); *Foxp3*^{gfp} *Aire*$^{KO/WT}$ ($n = 3$); *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/KO}$ ($n = 7$); *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/WT}$ ($n = 9$). Data are representative of >2 experiments. Two-tailed unpaired Mann–Whitney test. **d**, Analysis of tissue-specific autoantibodies in the serum of DKO mice. Sections of skin, intestine and eye from gender-matched *Rag1*-deficient mice were stained with serum IgG (>8 mice per group). DAPI, 4′,6-diamidino-2-phenylindole. **e**, **f**, Analysis of $Foxp3^+$ CD4 SP thymocytes (**e**) and peripheral T_{reg} cells (**f**) in *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/KO}$ mice. *Foxp3*^{gfp} *Aire*$^{KO/KO}$ ($n = 10$); *Foxp3*^{gfp} *Aire*$^{KO/WT}$ ($n = 6$); *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/KO}$ ($n = 11$); *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/WT}$ ($n = 11$). Two-tailed unpaired Mann–Whitney test.

the TCR repertoire of T_{reg} cells, which is essential for controlling pathogenic self-reactive T cells that escape negative selection.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 June; accepted 12 October 2015.
Published online 25 November 2015.

1. Klein, L., Kyewski, B., Allen, P. M. & Hogquist, K. A. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nature Rev. Immunol.* **14,** 377–391 (2014).
2. Josefowicz, S. Z., Lu, L. F. & Rudensky, A. Y. Regulatory T cells: mechanisms of differentiation and function. *Annu. Rev. Immunol.* **30,** 531–564 (2012).
3. Chandra, S. & Kronenberg, M. Activation and function of iNKT and MAIT cells. *Adv. Immunol.* **127,** 145–201 (2015).
4. Hsieh, C. S., Zheng, Y., Liang, Y., Fontenot, J. D. & Rudensky, A. Y. An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nature Immunol.* **7,** 401–410 (2006).
5. Hsieh, C. S. *et al.* Recognition of the peripheral self by naturally arising $CD25^+CD4^+$ T cell receptors. *Immunity* **21,** 267–277 (2004).
6. Lee, H. M., Bautista, J. L., Scott-Browne, J., Mohan, J. F. & Hsieh, C. S. A broad range of self-reactivity drives thymic regulatory T cell selection to limit responses to self. *Immunity* **37,** 475–486 (2012).
7. Levine, A. G., Arvey, A., Jin, W. & Rudensky, A. Y. Continuous requirement for the TCR in regulatory T cell function. *Nature Immunol.* **15,** 1070–1078 (2014).
8. Jordan, M. S. *et al.* Thymic selection of $CD4^+CD25^+$ regulatory T cells induced by an agonist self-peptide. *Nature Immunol.* **2,** 301–306 (2001).
9. Lafaille, J. J., Nagashima, K., Katsuki, M. & Tonegawa, S. High incidence of spontaneous autoimmune encephalomyelitis in immunodeficient anti-myelin basic protein T cell receptor transgenic mice. *Cell* **78,** 399–408 (1994).
10. Apostolou, I., Sarukhan, A., Klein, L. & von Boehmer, H. Origin of regulatory T cells with known specificity for antigen. *Nature Immunol.* **3,** 756–763 (2002).
11. Zheng, Y. *et al.* Role of conserved non-coding DNA elements in the *Foxp3* gene in regulatory T-cell fate. *Nature* **463,** 808–812 (2010).
12. Fontenot, J. D. *et al.* Regulatory T cell lineage specification by the forkhead transcription factor Foxp3. *Immunity* **22,** 329–341 (2005).
13. Setoguchi, R., Hori, S., Takahashi, T. & Sakaguchi, S. Homeostatic maintenance of natural *Foxp3*$^+$ $CD25^+$ $CD4^+$ regulatory T cells by interleukin (IL)-2 and induction of autoimmune disease by IL-2 neutralization. *J. Exp. Med.* **201,** 723–735 (2005).
14. Fontenot, J. D., Rasmussen, J. P., Gavin, M. A. & Rudensky, A. Y. A function for interleukin 2 in Foxp3-expressing regulatory T cells. *Nature Immunol.* **6,** 1142–1151 (2005).
15. Seddon, B., Legname, G., Tomlinson, P. & Zamoyska, R. Long-term survival but impaired homeostatic proliferation of naive T cells in the absence of $p56^{lck}$. *Science* **290,** 127–131 (2000).
16. Nicodeme, E. *et al.* Suppression of inflammation by a synthetic histone mimic. *Nature* **468,** 1119–1123 (2010).
17. Arpaia, N. *et al.* Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* **504,** 451–455 (2013).
18. Davie, J. R. Inhibition of histone deacetylase activity by butyrate. *J. Nutr.* **133,** 2485S–2493S (2003).
19. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504,** 446–450 (2013).
20. Moran, A. E. *et al.* T cell receptor signal strength in T_{reg} and iNKT cell development demonstrated by a novel fluorescent reporter mouse. *J. Exp. Med.* **208,** 1279–1289 (2011).
21. Zikherman, J., Parameswaran, R. & Weiss, A. Endogenous antigen tunes the responsiveness of naive B cells but not T cells. *Nature* **489,** 160–164 (2012).
22. Shugay, M. *et al.* Towards error-free profiling of immune repertoires. *Nature Methods* **11,** 653–655 (2014).
23. Shugay, M. *et al.* VDJtools: unifying post-analysis of T cell receptor repertoires. *PLOS Comput. Biol.* http://dx.doi.org./10.1371/journal.pcbi.1004503 (in the press).
24. Košmrlj, A., Jha, A. K., Huseby, E. S., Kardar, M. & Chakraborty, A. K. How the thymus designs antigen-specific and self-tolerant T cell receptor sequences. *Proc. Natl Acad. Sci. USA* **105,** 16671–16676 (2008).
25. Yang, S., Fujikado, N., Kolodin, D., Benoist, C. & Mathis, D. Immune tolerance. Regulatory T cells generated early in life play a distinct role in maintaining self-tolerance. *Science* **348,** 589–594 (2015).
26. Klein, L., Hinterberger, M., Wirnsberger, G. & Kyewski, B. Antigen presentation in the thymus for positive selection and central tolerance induction. *Nature Rev. Immunol.* **9,** 833–844 (2009).
27. Liston, A., Lesage, S., Wilson, J., Peltonen, L. & Goodnow, C. C. Aire regulates negative selection of organ-specific T cells. *Nature Immunol.* **4,** 350–354 (2003).
28. Giraud, M. *et al.* Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. *Proc. Natl Acad. Sci. USA* **109,** 535–540 (2012).
29. Malchow, S. *et al.* Aire-dependent thymic development of tumor-associated regulatory T cells. *Science* **339,** 1219–1224 (2013).
30. Yates, A. J. Theories and quantification of thymic selection. *Front. Immunol.* **5,** 13 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Bos, A. Arvey, C. Konopacki, G. Gasteiger, S. Lee, T. Chinen and K. Wu for technical assistance, CKP IBCH for equipment, and R. Prinjha for providing iBET. Y.F. was supported by a Postdoctoral Fellowship of the Cancer Research Institute. This study was supported by NIH grants R37 AI034206 and U01 HG007893, Cancer Center Support Grant P30 CA008748, and the Howard Hughes Medical Institute (A.Y.R.). M.S., E.V.P. and D.M.C. were supported by MCB program RAS and RFBR grants 14-04-01247 and 15-34-21052.

Author Contributions Y.F. and A.R. conceived and designed the study. Y.F. performed animal and *in vitro* studies, flow cytometric, TCR sequencing and gene expression analyses. J.v.d.V. analysed the epigenetic modifications of CNS3 and how they affect Foxp3 transcriptional regulation. M.S., E.V.P. and D.M.C. analysed TCR sequencing data. H.U.O. and C.S.L. analysed RNA sequencing data. B.E.H. performed serum Ig isotype analysis. S.D. and S.H. participated in phenotypic analysis of mice. S.H. generated the Cre retroviral construct. B.M. and S.D. participated in optimizing TCR sequencing protocol. P.T. analysed histopathology. Y.F. and A.Y.R. wrote the manuscript.

Author Information All RNA and TCR sequencing data have been deposited in the Gene Expression Omnibus under accession numbers GSE71309 and GSE71162, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.Y.R. (rudenska@mskcc.org).

METHODS

No statistical methods were used to determine sample size.

Mice. *Foxp3*$^{CNS3\text{-}fl\text{-}gfp}$ mice were generated using ES cell line CY2.4 (C56BL/6) as previously described[11]. *Cd4*^{Cre}, *Lck*^{Cre}, *Ubc*$^{Cre\text{-}ERT2}$ and Rosa26-stop-YFP (R26Y) mice were obtained from the Jackson Laboratories. DO11.10 TCRβ transgenic and *Aire*-knockout mice were provided by P. Marrack, and D. Mathis and C. Benoist, respectively. Heterozygous females carrying *Foxp3*$^{\Delta CNS3\text{-}gfp}$ and *Foxp3*^{gfp} were crossed with B6 males to generate hemizygous *Foxp3*$^{\Delta CNS3\text{-}gfp}$ and wild-type *Foxp3*^{gfp} littermates. *Foxp3*^{DTR}, *Foxp3*-null, *Rag1*$^{-/-}$, CD45.1^{+} *Foxp3*^{gfp} and *Tcrb*$^{-/-}$ *Tcrd*$^{-/-}$ mice were maintained in our animal facility. To study the genetic interactions between CNS3 and *Aire*, heterozygous females of *Foxp3*$^{\Delta CNS3\text{-}gfp/gfp}$ were first crossed with *Aire*$^{KO/WT}$, and F_1 harbouring *Aire*$^{KO/WT}$ and *Foxp3*$^{\Delta CNS3\text{-}gfp}$ or *Foxp3*^{gfp} were then intercrossed to generate *Aire*$^{KO/KO}$ or *Aire*$^{KO/WT}$ mice carrying *Foxp3*$^{\Delta CNS3\text{-}gfp}$ or *Foxp3*^{gfp}. To examine TCR diversity with restricted repertoire, *Foxp3*$^{\Delta CNS3\text{-}gfp/gfp}$ heterozygous females were crossed to the DO11.10 TCRβ transgenic and *Tcra*$^{-/+}$ males. F_1 males of *Foxp3*$^{\Delta CNS3\text{-}gfp}$ or *Foxp3*^{gfp} mice carrying the DO11.10 TCRβ transgene and *Tcra*$^{-/+}$ were used for T-cell isolation and TCR sequencing. To induce deletion of CNS3 *in vivo*, tamoxifen solution (40 mg ml^{-1} in olive oil) was administered by gavage to *Ubc*$^{Cre\text{-}ERT2}$ *Foxp3*$^{CNS3\text{-}fl\text{-}gfp}$ R26Y mice more than 3 days before lymphocyte isolation.

All mice were maintained in the MSKCC animal facility under SPF conditions, and the experiments were approved by the Institutional Review Board (IACUC 08-10-023). The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Statistical analysis. Statistical tests were performed with Prism (GraphPad), Excel (Microsoft) or R statistical environment. Box-and-whisker plots show minimum, maximum, first and third quartiles and median.

Cell culture. For *in vitro* T_{reg} cell differentiation, naive CD4^{+} T cells (GFP^{-}CD25^{-}CD44^{lo}CD62L^{hi}) or mature CD4^{+}CD8^{-} SP (TCRβ^{hi}GFP^{-}CD25^{-}CD62L^{hi}CD69^{lo}) T cells were sorted from *Foxp3*^{gfp}, *Foxp3*$^{\Delta CNS3\text{-}gfp}$ or *Foxp3*$^{CNS3\text{-}fl\text{-}gfp}$ mice after the enrichment of CD4^{+} T cells or depletion of CD8^{+} T cells using Dynabeads FlowComp Mouse CD4 or CD8 kits, respectively (Life Technologies), and then cultured with lethally irradiated (20 Gy) antigen-presenting cells (splenocytes depleted of T cells with Dynabeads FlowComp Mouse CD90.2 kit, Life Technologies) or on plates pre-coated with CD3 and CD28 antibodies in RPMI1640 supplemented with 10% fetal bovine serum (FBS), 2 mM L-glutamine, 1 mM sodium pyruvate, 10 mM HEPES, 2×10^{-5} M 2-mercaptoethanol, 100 U ml^{-1} penicillin, 100 mg ml^{-1} streptomycin, 500 U ml^{-1} IL-2 and 1 ng ml^{-1} TGFβ. Sodium butyrate (water solution) or iBET solution in dimethylsulfoxide (a gift from R. Prinjha) was added to the culture to block histone deacetylase or bromodomain-containing proteins, respectively. T_{reg} cells were sorted on the basis of *Foxp3*^{gfp} reporter expression. Assessment of the stability of Foxp3 expression *in vitro* was performed as previously described[31]. Briefly, T_{reg} cells were activated in culture in the presence of CD3 and CD28 antibody-coated beads (Life Technologies) with the following recombinant pro-inflammatory cytokines: IL-2 (250 U ml^{-1}), IL-4 (20 ng ml^{-1}), IL-6 (10 ng ml^{-1}), IFNγ (100 ng ml^{-1}) and IL-12 (20 ng ml^{-1}).

***In vivo* suppression assay.** To assess T_{reg}-cell suppressor capacity *in vivo* we conducted adoptive T-cell transfers into T-cell-deficient recipients as previously described[31]. Briefly, $\sim 2.5 \times 10^6$–3.0×10^6 Foxp3^{-}CD4^{+} and/or CD8^{+} T cells isolated from *Foxp3*-null or *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/KO}$ mice were transferred to congenic and gender-matched *Tcrb*$^{-/-}$ *Tcrd*$^{-/-}$ recipients alone or at a 10:1 ratio with sorted T_{reg} cells from *Foxp3*^{gfp} or *Foxp3*$^{\Delta CNS3\text{-}gfp}$ littermates. Similar numbers of effector T cells and T_{reg} cells were used for *in vivo* evaluation of T_{reg} suppressor function after acute ablation of CNS3. Recipient mice were monitored for body weight change regularly and lymphocytes were analysed by flow cytometry at least 4 weeks after the transfer.

T_{reg} cell homeostatic proliferation in lymphopenic mice. Around 8–10 weeks after bone marrow reconstitution of CD45.1^{+} *Foxp3*^{gfp} and CD45.2^{+} *Foxp3*$^{\Delta CNS3\text{-}gfp}$ in *Tcrb*$^{-/-}$ *Tcrd*$^{-/-}$ recipients, CD45.1^{+} and CD45.2^{+} T_{reg} cells (CD4^{+}GFP^{+}) were sorted, mixed at a 1:1 ratio and co-transferred into *Tcrb*$^{-/-}$ *Tcrd*$^{-/-}$ male mice with tenfold naive CD4^{+} T cells (CD25^{-}CD44^{lo}CD62L^{hi}) isolated from wild-type CD45.2^{+} B6 males. To block TCR stimulation by pMHC-II complexes, 0.5 mg of I-Ab-specific monoclonal antibody Y3P (IgG2a) or control IgG2a (Bio X Cell) was injected intravenously every other day before and after T-cell transfer[32,33]. The lymphocyte subsets were analysed by flow cytometry 9 days later.

Flow cytometric analyses and tissue lymphocyte preparation. Tissue lymphocytes were prepared as previously described[31]. The following fluorophore-conjugated antibodies were used for cell-surface staining: CD4 (RM4-5, eBioscience), CD8 (5H10, Life Technologies), CD25 (PC61.5, eBioscience), CD3e (145-2C11, eBioscience), CD44 (IM7, eBioscience), CD62L (MEL-14, eBioscience), CTLA4 (UC10-4B9, eBioscience), TCRβ (BioLegend), CD45.1 (A20, eBioscience) and CD45.2 (104, eBioscience). Antibodies used for intracellular staining were: Foxp3 (FJK-16 s, eBioscience), Ki-67 (B56, BD Biosciences), IL-17 (eBio17B7, eBioscience), IFNγ (XMG1.2, eBioscience) and IL-2 (JES6-5H4, eBioscience). To stain endogenous Nur77, cells were incubated with rabbit-anti-Nur77 antibody (Cell Signaling) after fixation and permeabilization with a Foxp3/transcription-factor-staining buffer set (eBioscience), followed by phycoerythrin-conjugated donkey anti-rabbit antibody (eBioscience). For the flow cytometric analysis of cytokine production, lymphocytes were first stimulated *in vitro* with 10 mg ml^{-1} of CD3 antibody in the presence of monensin (BD Biosciences) at 37 °C for 5 h, then stained with antibodies against indicated cell-surface markers followed by staining of cytokines with an intracellular staining kit (BD Biosciences). All flow cytometric analyses were performed using live-cell gate defined as negative by staining with the LIVE/DEAD Fixable Dead Cell Stain Kit (Life Technologies). Flow cytometric analysis was performed with FlowJo (Treestar).

Retroviral transduction. The Cre coding region was subcloned into MigR1-IRES-Thy1.1 vector (A. Levine, unpublished data) to generate MigR1-Cre-IRES-Thy1.1. Retroviral packaging with regular Phoenix-ECO cells and transduction of T_{reg} cells were performed following standard protocols[5].

Autoantibody profiling using autoantigen microarrays. Analysis of autoantibody reactivity against a panel of 95 autoantigens was conducted using the autoantigen microarrays developed by University of Texas Southwestern Medical Center[34]. Briefly, serum samples pretreated with DNase-I and diluted at 1:50 were incubated with the auto-antigen arrays. After a second incubation with Cy3-conjugated anti-mouse IgG, the arrays were scanned with a Genepix 4200A scanner (Molecular Device). The fluorescent signals for individual autoantigens were extracted from the resulting images with Genepix Pro 6.0 (Molecular Devices), followed sequentially by subtraction of local background, average of duplicates, normalization with total IgG, and subtraction of a negative PBS control.

TCR sequencing and data analysis. *Cell isolation and RNA extraction.* Lymphocytes were collected from the peripheral lymphoid organs or thymi of 6–8-week-old male *Foxp3*^{gfp} or *Foxp3*$^{\Delta CNS3\text{-}gfp}$*Tcra*$^{-/+}$ littermates bearing the DO11.10 TCRβ transgene, and were enriched for CD4^{+} T cells (Dynabeads FlowComp Mouse CD4 kit, Life Technologies) or depleted of CD8^{+} T cells (Dynabeads FlowComp Mouse CD8 kit, Life Technologies), respectively, and T_{reg} cells (CD4^{+}GFP^{+}), mature Foxp3^{-}CD4 SP thymocytes (CD4^{+}CD8^{-}GFP^{-}CD25^{-}CD62L^{hi}CD69^{lo}), peripheral naive (CD4^{+}GFP^{-}CD25^{-}CD44^{lo}CD62L^{hi}) and effector (CD4^{+}GFP^{-}CD44^{hi}CD62L^{lo}) CD4^{+} T cells were isolated using a FACSAria II sorter (BD) gated on TCR-Vβ8^{hi}. Extraction of total RNA from TRIzol-preserved cell lysates was performed according to the manufacturer's instructions (Life Technologies). mRNA was purified from total RNA with Dynabeads mRNA DIRECT Kit (Life Technologies) and used for reverse transcription.

cDNA synthesis. To maximize the priming efficiency of reverse transcription, a mixture of oligo(dT)$_{24}$ and eight DNA oligonucleotides corresponding to the mouse TCRα constant region was used. The oligonucleotides used in this study were synthesized by Integrated DNA Technologies, Inc.

TRAC_RT1: 5'-CTCAGCGTCATGAGCAGGTTAAAT-3', TRAC_RT2: 5'-CAGGAGGATTCGGAGTCCCATAA-3', TRAC_RT3: 5'-TTTTACAACATTCTCCAAGA-3', TRAC_RT4: 5'-TTCTGAATCACCTTTAATGA-3', TRAC_RT5: 5'-ATGAGATAATTTCTACACCT-3', TRAC_RT6: 5'-TTTGGCTTGAAGAAGGAGCG-3', TRAC_RT7: 5'-TTCAAAGCTTTTCTCAGTCA-3', TRAC_RT9: 5'-TGGTCTCTTTGAAGATATCT-3'.

To label the 5' end of TCRα mRNA, a DNA–RNA hybrid oligonucleotide with 12 random nucleotides serving as barcodes to tag individual mRNA molecules was synthesized as previously reported[35].

Hybrid oligonucleotide: AAGCAGTGGTATCAACGCAGAGUNNNNUNNNNUNNNNUCTTrGrGrGrGrG (r, ribonucleotide). cDNA was synthesized in SMARTScribe reverse-transcription buffer (Clontech) with 1.0 μM each of reverse transcription oligonucleotide, 0.5 mM of each dNTP, 5.0 mM of dithiothreitol (DTT), 2.0 U μl^{-1} recombinant RNase inhibitor (Takara), 1 μM hybrid oligonucleotide, 1 M betaine (Affymetrix), 6 mM $MgCl_2$ and 5 U μl^{-1} SMARTScribe reverse transcriptase by incubating at 42 °C for 90 min, followed by 10 cycles of incubation at 50 °C for 2 min, 42 °C for 2 min, and then one step of incubation at 70 °C for 15 min. After removal of hybrid oligonucleotide with Uracil-DNA Glycosylase (New England BioLabs), cDNA was purified with Agencourt AMPure XP beads (Beckman Coulter) according the manufacturer's manual.

Sequencing library preparation. Purified cDNA was used as templates for a four-step PCR amplification, in which sequencing adaptors and sample indices were introduced. The first PCR reaction was performed with purified cDNA, 0.2 μM universal primer (5'-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3', Clontech), 0.2 μM TRAC

reverse primer 8 (5'-TTTTGTCAGTGATGAACGTT-3'), 0.2 mM each dNTP, 1.5 mM $MgCl_2$ and 0.02 U μl^{-1} KOD Hot Start DNA Polymerase (EMD Millipore). PCR parameters were as follows: initial denature at 95 °C for 2 min; 10 cycles of 95 °C for 20 s, 70 °C for 10 s with an increment of −1 °C per cycle, and 70 °C for 30 s; 15 cycles of 95 °C for 20 s, 60 °C for 10 s and 70 °C for 30 s; and final cycle at 70 °C for 3.5 min. Amplified DNA was purified with Agencourt AMPure XP magnetic beads for the subsequent reaction. The second PCR reaction used the same reactants except that the reverse primer was replaced by a nested primer (5'-CAATTGCACCCTTACCACGACAGTCTGGTACACAG CAGGTTCTGGGTTCTGGA-3'). Cycling parameters were: 95 °C for 2 min; 6 cycles of 95 °C for 20 s, 60 °C 10 s and 70 °C 30 s; and a final cycle at 70 °C for 3.5 min. DNA from individual samples was extracted with Agencourt AMPure XP magnetic beads and used for the third round of amplification with 5RACE TCR forward primer (5'-AATGATACGGCGACCACCGAGATCTACACCTA ATACGACTCACTATAGGGC-3') and indexed reverse primer (5'-CAAGCA GAAGACGGCATACGAGATXXXXXXAGTCAGTCAGCCCAATTGCACC CTTACCACGA-3', XXXXXX for 6-nucleotide barcode). The cycling parameters were: 95 °C for 2 min; 6 cycles at 95 °C for 20 s, 55 °C for 10 s and 70 °C for 30 s; and a final cycle at 70 °C for 3.5 min. The PCR products were purified with Agencourt AMPure XP magnetic beads and used for the fourth PCR amplification with primers P1 (5'-AATGATACGGCGACCACCGAG-3') and P2 (5'-CAAGCAGAAGACGGCATACGA-3'), and the following cycling parameters: 95 °C for 2 min; 5 cycles at 95 °C for 20 s, 57 °C for 10 s and 70 °C for 30 s; and a final cycle at 70 °C for 3.5 min. The final PCR products were separated by agarose gel electrophoresis and a single band around 600 base-pairs was cut and extracted with Gel Extraction and PCR Clean-Up kits (Takara).

High-throughput sequencing. Samples were quantified with Kapa Library Quantification kits (Kapa Biosystems) and sequenced on a MiSeq sequencer (Illumina) using 200 cycles of read 1, 6 cycles of index read and 200 cycles of read 2 with the following customized primers: read 1: 5'-CTAATACGA CTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3'; index read: 5'-TCGTGGTAAGGGTGCAATTGGGCTGACTGACT-3'; read 2: 5'-AGTCAGTCAGCCCAATTGCACCCTTACCACGA-3'.

Data analysis. Barcoded sequencing data were analysed with MIGEC software[22]. Briefly, unique molecular identifier sequences were extracted from raw sequencing data (read 1) with MIGEC/Checkout routine. Reads ($\geq$5) bearing the same unique molecular identifier were grouped and assembled to generate consensus sequences with MIGEC/Assemble. Variable (V) and joining (J) segment mapping, CDR3 extraction, and error correction were performed with MIGEC/CdrBlast as previously described[22], which eliminates PCR and sequencing errors, as well as normalizes the output data as cDNA counts that represent the TCR clonotypes in a population[36].

Comparison of TCRα repertoires between CNS3-deficient and -sufficient mice at protein level was evaluated using VDJtools post-analysis framework (https://github.com/mikessh/vdjtools)[23]. Pearson correlation of clonotype frequencies for the shared TCR clones was used for the generation of the dendrogram. Clonal diversities of TCRα repertoires were evaluated using inverse Simpson index computed separately for individual samples after downsampling the repertoires to the size of the smallest sample from the same organ. Similar downsampling strategy, not weighted by clonotype frequencies, was used to compute the average size of added nucleotides in CDR3. A mathematical model[24] was used to assess the strength of CDR3 amino acid interactions with pMHC complexes. Numbers of strongly interacting amino acid residues (LFIMVWCY) were calculated for the V-segment part of TCRα CDR3 and V–J segment junction. Those numbers were then weighted by the corresponding clonotype frequencies and the resulting sums were used for the comparisons between samples.

RNA sequencing and data analysis. Mature $Foxp3^-CD4^+CD8^-$ SP ($TCR\beta^+GFP^-CD62L^{hi}CD69^{lo}$) thymocytes, $Foxp3^+$ CD4 SP thymocytes (thymic T_{reg} cells), peripheral resting ($CD44^{lo}CD62L^{hi}$) and activated ($CD44^{hi}CD62L^{lo}$) T_{reg} cells were FACS-sorted from ~6–8-week-old male *Foxp3*^{gfp} and *Foxp3*$^{\Delta CNS3\text{-}gfp}$ littermates. RNA was extracted and cDNA libraries were generated after SMART amplification (Clontech). Libraries were sequenced using a HiSeq 2000 platform (Illumina) according to a standard paired-end protocol. Reads were first processed with Trimmomatic[37] to remove TruSeq adaptor sequences and bases with quality scores below 20, and reads with less than 30 remaining bases were discarded. Trimmed reads were then aligned to mm10 mouse genome with the STAR spliced-read aligner[38]. For each gene from the RefSeq annotations, the number of uniquely mapped reads overlapping with the exons was counted with HTSeq (http://www-huber.embl.de/users/anders/HTSeq/). Genes with fewer than 50 read counts were considered as not expressed and filtered out. Principal component analysis (PCA) was performed ($n = 11{,}962$) for clustering gene expression. Differential gene expression was estimated using DESeq package[39]. To determine activation-related transcriptional signatures in T_{reg} cells, the differences between read counts of peripheral activated versus resting T_{reg} cells from wild-type *Foxp3*^{gfp} mice were evaluated by fold-change and Benjamini–Hochberg corrected *P* values (false discovery rate < 0.001) (Supplementary Data 1 and 2). For gene expression comparisons, previously published transcriptional signatures of TCR-dependent genes in T_{reg} cells were used[7]. The distribution of gene expression changes is shown for transcriptional signature genes and the rest of all expressed genes. One-tailed Kolmogorov–Smirnov test is used to determine the significance between the distributions of signature genes and the rest of expressed genes.

Chromatin immunoprecipitation. We cross-linked 1×10^6 cells with 1% formaldehyde for 5 min at room temperature. Cross-linked cells were lysed and nuclei were resuspended in 250 μl nuclear lysis buffer containing 1% SDS. Chromatin input samples were prepared by sonication of cross-linked nuclear lysates. For histone ChIPs, nuclear lysates were subjected to micrococcal nuclease (MNase) digestion before sonication. Nuclei were resuspended in 100 μl MNase (New England Biolabs) at 12,000 U ml^{-1} for 1 min at 37 °C. The reaction was stopped by addition of 10 μl of 0.5 M EDTA. Chromatin input samples were incubated overnight at 4 °C with antibodies against H3K4me1 (Abcam), H3K4me3 (Millipore) or H3K27ac (Abcam), and precipitated for 90 min at 4 °C using protein A Dynabeads (Life Technologies). After thorough washing, bead-bound chromatin was subjected to proteinase K digestion and decrosslinking overnight at 65 °C. DNA fragments were isolated using a Qiagen PCR purification kit. Relative abundance of precipitated DNA fragments was analysed by qPCR using Power SYBR Green PCR Master Mix (Applied Biosystems). The following primers were used for qPCR: *Gm5069*: forward: 5'-TAAGCAATTGGTGGTGCAGGATGC-3', reverse: 5'-AAAGGGTCATCATCTCCGTCCGTT-3'; *Hspa2*: forward: 5'-TC GTGGAGAGTTGTGAGAAGCGA-3', reverse: 5'-AACGTTAGGACGAAA GCGTCAGGA-3'; *Hsp90ab*: forward: 5'-TTACCTTGACGGGAAAGCCG AGTA-3', reverse: 5'-TTCGGGAGCTCTCTTGAGTCACC-3'; *Rpl30*: forward: 5'-TCGGCTTCACTCACCGTCTTCTTT-3', reverse: 5'-TG TCCTCTGTGTATGCTAGGTTGG-3'; *Foxp3* promoter: forward: 5'-TAATGTGGCAGTTTCCCACAAGCC-3', reverse: 5'-AATACCTC TCTGCCACTT TCGCCA-3'; CNS1: forward: 5'-AGACTGTCTGGA ACAACCTAGCCT-3', reverse: 5'-TGGAGGTACAGAGAGGTTAAGAGCCT-3'; CNS2: forward: 5'-ATCTGGCCAAGTTCAGGTTGTGAC-3', reverse: 5'-GGGCGTTCCTGTTTGACTGTTTCT-3'; CNS3: forward: 5'-TCTCC AGGCTTCAGAGATTCAAGG-3', reverse: 5'-ACAGTGGGATGAGG ATACATGGCT-3'.

Relative enrichment was calculated by normalizing to background binding to the control region (*Gm5069*).

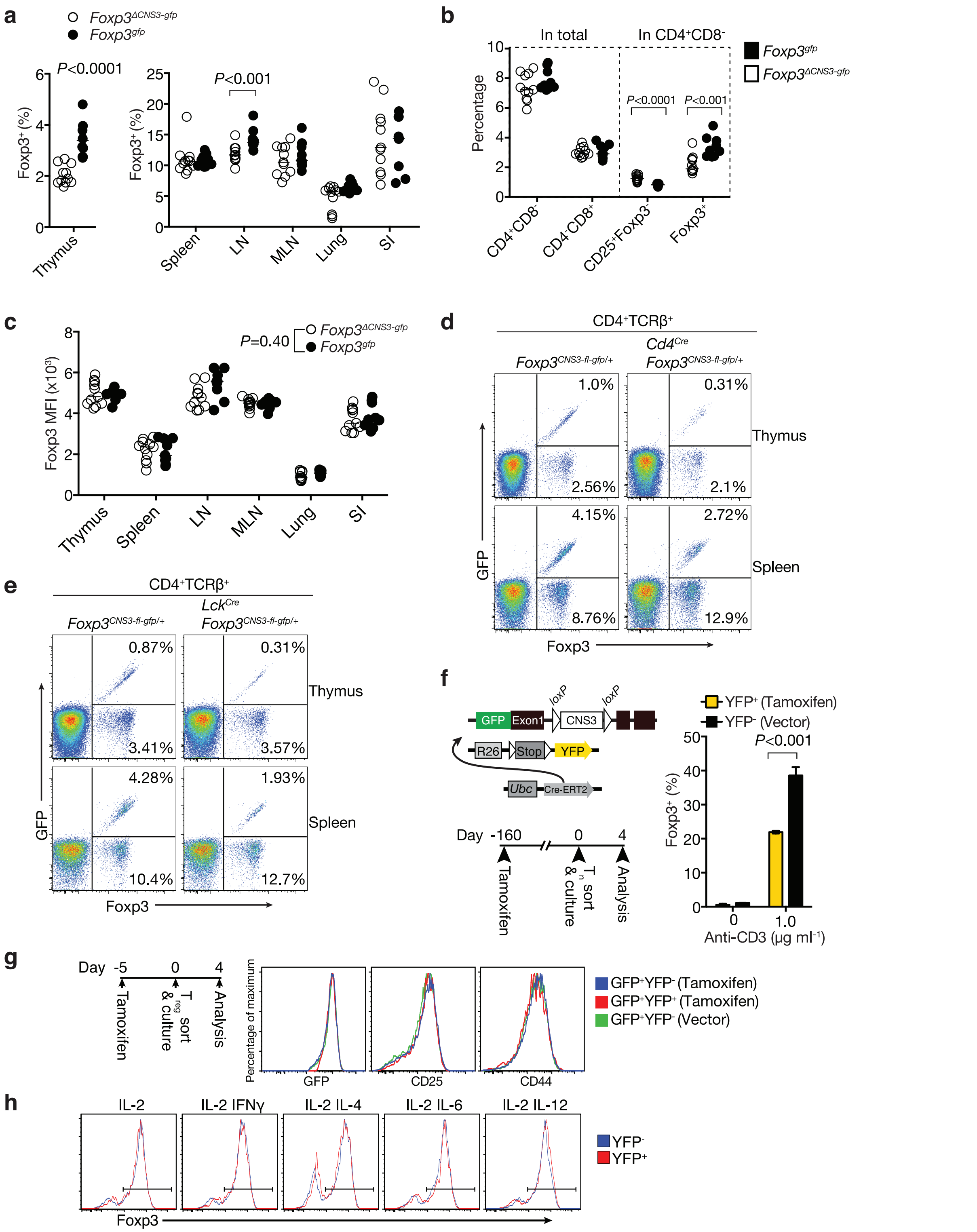
Ig isotype ELISA and immunofluorescence staining. Quantification of serum Ig isotypes was performed by ELISA as previously described[40]. Tissue sections from gender matched *Rag1*$^{-/-}$ mice were used to detect mouse autoantibodies. Briefly, organs from the *Rag1*$^{-/-}$ mice were dissected, fixed with neutral buffered formalin, embedded with paraffin and sectioned. After deparaffinization with EZPrep buffer (Ventana Medical Systems) and antigen retrieval with cell conditioning solution (Ventana Medical Systems) the sections were blocked for 30 min with Background Buster solution (Innovex), followed by avidin/biotin blocking for 8 min, mouse serum (1:50 dilution) incubation for 5 h and biotinylated horse anti-mouse IgG (Vector Labs) incubation for 1 h. The detection was performed with streptavidin–horseradish peroxidase (Ventana Medical Systems) followed by incubation with Tyramide Alexa Fluor 488 (Invitrogen). The slides were then counterstained with DAPI (Sigma Aldrich) for 10 min, mounted, scanned with a Mirax scanner and visualized with Pannoramic Viewer (3DHISTECH). Scanned images were scored and representative snapshots were processed with Photoshop (Adobe) to switch the green and red channels for presentation purpose.

Generation of mixed bone marrow chimaeras. Mixed bone marrow chimaeras were generated as previously described[31]. Briefly, recipient mice were irradiated (9.5 Gy) 24 h before intravenous injection of 10×10^6 bone marrow cells from $CD45.1^+$ *Foxp3*^{gfp} and $CD45.2^+$ *Foxp3*$^{\Delta CNS3\text{-}gfp}$ mixed at a 1:1 ratio. After bone marrow transfer, the recipient mice were administrated with 2 mg ml^{-1} neomycin in drinking water for 3 weeks and analysed 8–10 weeks later.

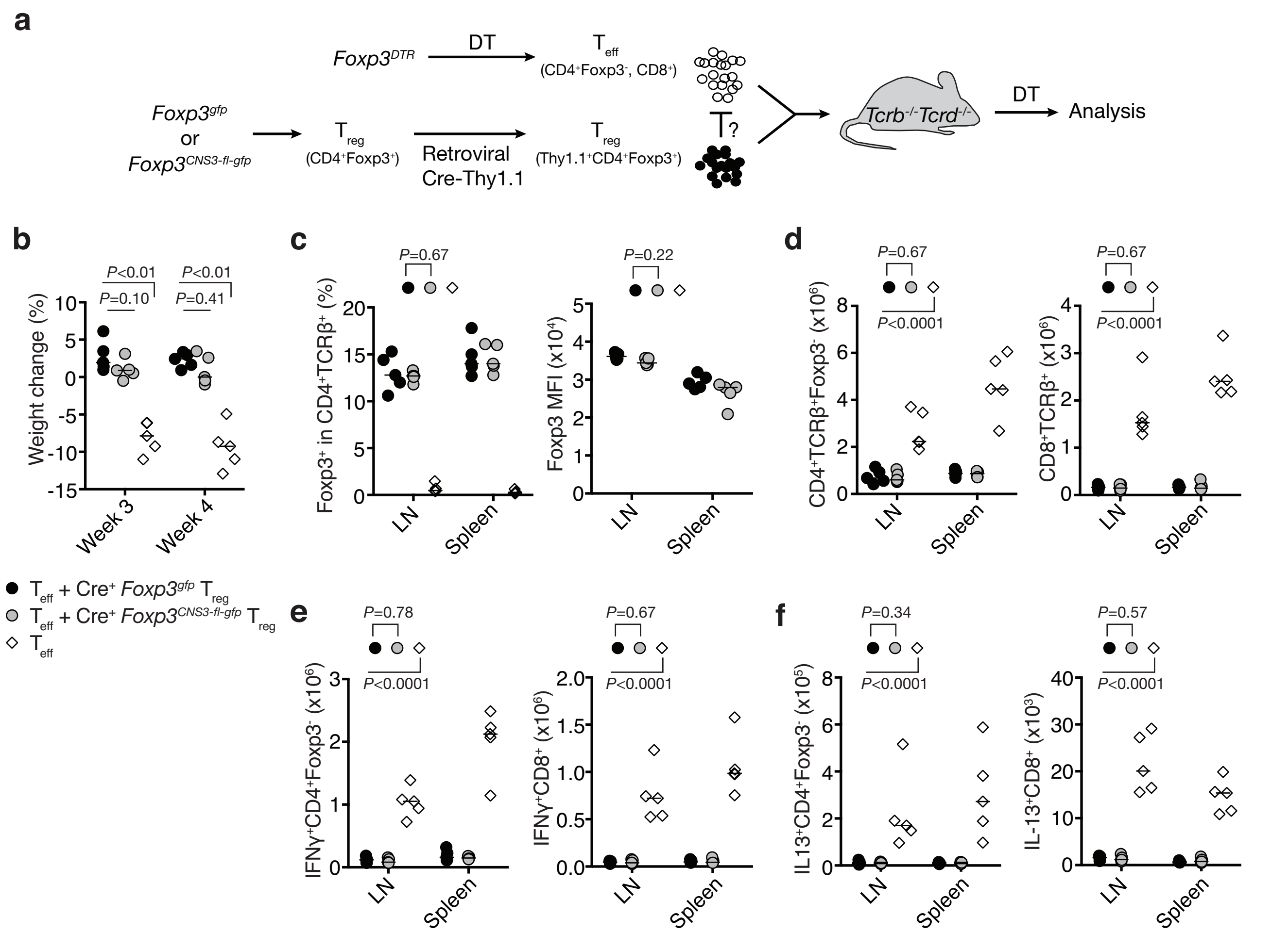
Histological analysis. Tissue samples were fixed in 10% neutral buffered formalin and processed for haematoxylin and eosin staining. Stained slides were scored for tissue inflammation as previously described[41].

Experimental autoimmune encephalomyelitis induction. Experimental autoimmune encephalomyelitis was induced by immunization with myelin oligodendrocyte glycoprotein peptide 35-55 (MOG35-55, GenScript) in complete Freund's adjuvant (CFA, Sigma) and mice were monitored for disease as previously described[42].

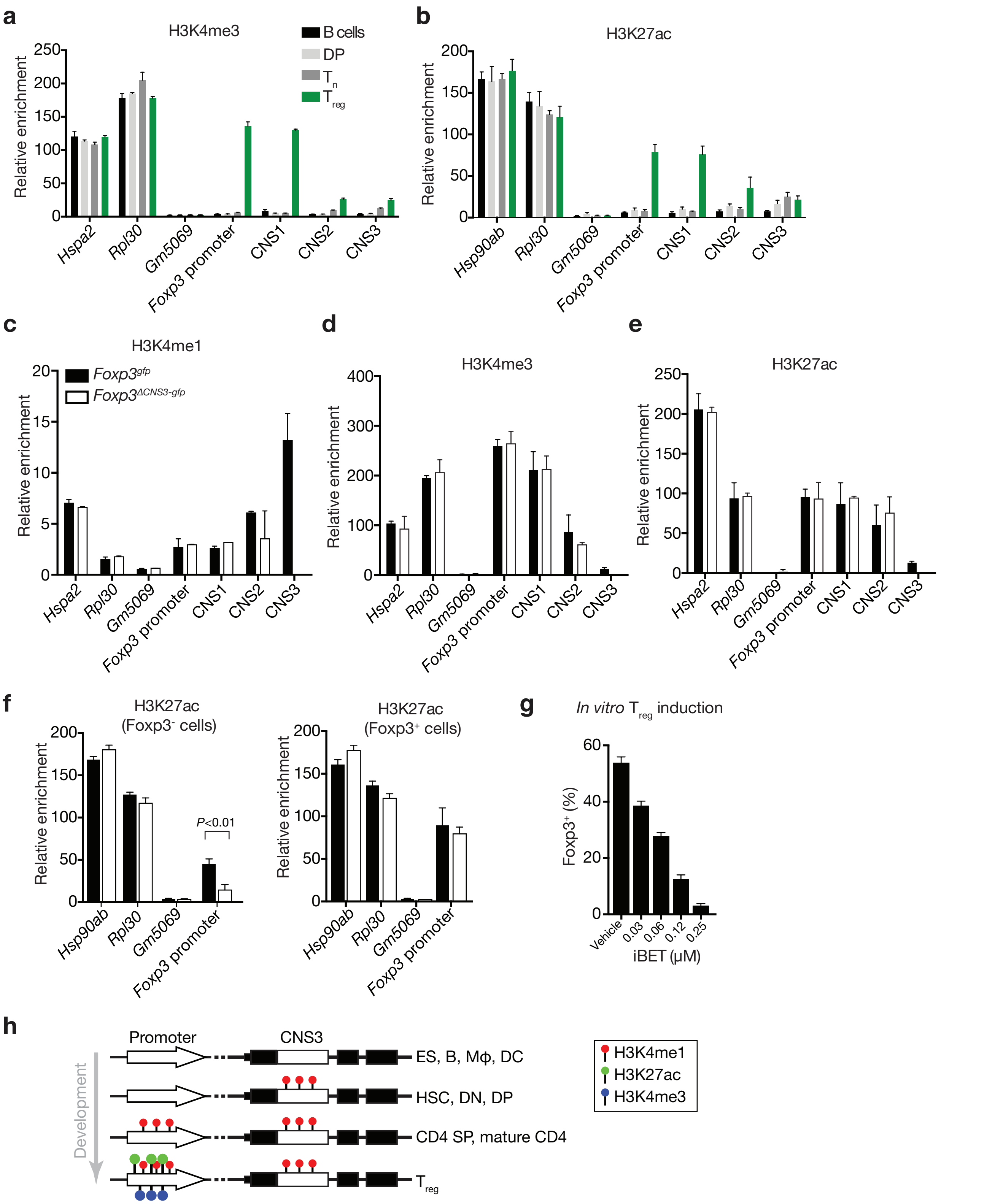
31. Feng, Y. *et al.* Control of the inheritance of regulatory T cell identity by a *cis* element in the *Foxp3* locus. *Cell* **158,** 749–763 (2014).
32. Janeway, C. A. Jr *et al.* Monoclonal antibodies specific for Ia glycoproteins raised by immunization with activated T cells: possible role of T cellbound Ia antigens as targets of immunoregulatory T cells. *J. Immunol.* **132,** 662–667 (1984).
33. Stefanová, I., Dorfman, J. R. & Germain, R. N. Self-recognition promotes the foreign antigen sensitivity of naive T lymphocytes. *Nature* **420,** 429–434 (2002).
34. Zhen, Q. L. *et al.* Identification of autoantibody clusters that best predict lupus disease activity using glomerular proteome arrays. *J. Clin. Invest.* **115,** 3428–3439 (2005).
35. Egorov, E. S. *et al.* Quantitative profiling of immune repertoires for minor lymphocyte counts using unique molecular identifiers. *J. Immunol.* **194,** 6155–6163 (2015).
36. Britanova, O. V. *et al.* Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. *J. Immunol.* **192,** 2689–2698 (2014).
37. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
38. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
39. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).
40. Zheng, Y. *et al.* Regulatory T-cell suppressor program co-opts transcription factor IRF4 to control T_H2 responses. *Nature* **458,** 351–356 (2009).
41. Chaudhry, A. *et al.* $CD4^+$ regulatory T cells control T_H17 responses in a Stat3-dependent manner. *Science* **326,** 986–991 (2009).
42. Stromnes, I. M. & Goverman, J. M. Active induction of experimental allergic encephalomyelitis. *Nature Protocols* **1,** 1810–1819 (2006).



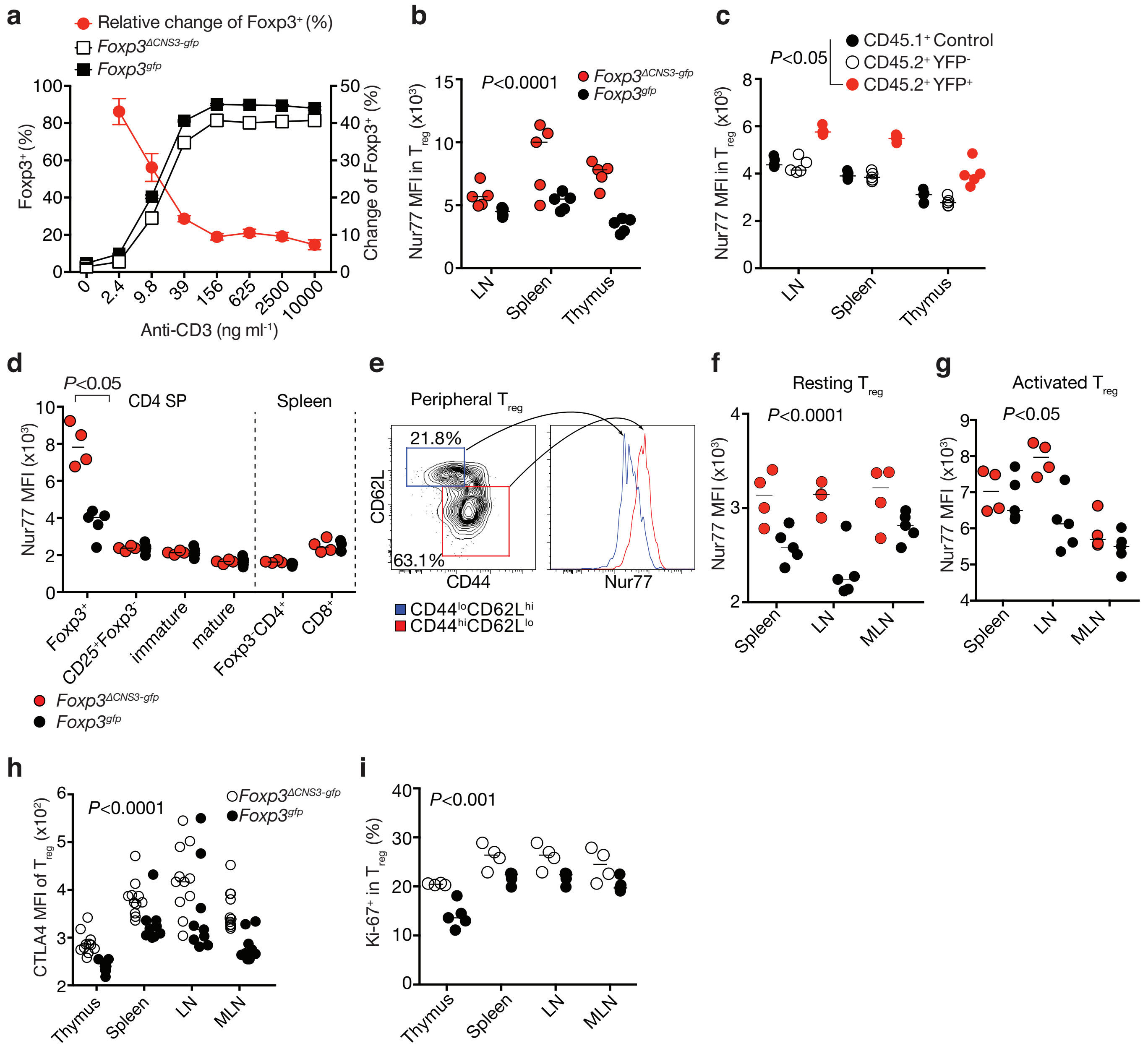
Extended Data Figure 1 | CNS3 is required in precursor cells for optimal T_{reg} cell differentiation. a, Diminished numbers of thymic T_{reg} cells in 6–8-week-old CNS3-deficient mice. Two-tailed Mann–Whitney test. The data show individual mice and median, and represent 1 of >2 independent experiments. *Foxp3*^{gfp} ($n = 9$); *Foxp3*$^{\Delta CNS3\text{-}gfp}$ ($n = 11$). SI, small intestine. **b**, **c**, Flow cytometric analysis of CD4 and CD8 SP thymocyte subsets, including thymic T_{reg} precursor ($CD4^+CD25^+Foxp3^-$) cells (**b**) and Foxp3 expression (**c**) in 6–8-week-old *Foxp3*$^{\Delta CNS3\text{-}gfp}$ mice ($n = 11$) and *Foxp3*^{gfp} ($n = 9$) littermates. Unpaired Mann–Whitney test. **d**, **e**, CNS3-dependent T_{reg} cell differentiation in heterozygous *Foxp3*$^{CNS3\text{-}fl\text{-}gfp/+}$ and *Cd4*^{Cre} *Foxp3*$^{CNS3\text{-}fl\text{-}gfp/+}$ (**d**), or *Foxp3*$^{CNS3\text{-}fl\text{-}gfp/+}$ and *Lck*^{Cre} *Foxp3*$^{CNS3\text{-}fl\text{-}gfp/+}$ females (**e**). GFP^+ and GFP^- T_{reg} cells in these mice express *Foxp3*$^{CNS3\text{-}fl\text{-}gfp}$ or wild-type *Foxp3*$^+$ alleles, respectively. The data represent 1 of >2 independent experiments ($n \geq 3$ mice per group). **f**, Acute ablation of CNS3 impairs T_{reg} induction *in vitro*. Yellow fluorescent protein $(YFP)^+$ (tamoxifen treated) or YFP^- (vector control) naive $CD4^+$ T cells from *Ubc*$^{Cre\text{-}ERT2}$ *Foxp3*$^{CNS3\text{-}fl\text{-}gfp}$ R26Y males were cultured under *in vitro* T_{reg} induction conditions. The data show mean ± s.e.m. of triplicate cultures and represent 1 of 2 independent experiments. Two-tailed unpaired *t*-test. **g**, **h**, Acute ablation of CNS3 in differentiated T_{reg} cells does not affect Foxp3 expression level on a per cell basis or the stability of mature T_{reg} cells. **g**, Expression of Foxp3, CD25 and CD44 in T_{reg} cells on day 4 after tamoxifen treatment. **h**, YFP^+ and YFP^- T_{reg} cells from tamoxifen-treated *Ubc*$^{Cre\text{-}ERT2}$ *Foxp3*$^{CNS3\text{-}fl\text{-}gfp}$ R26Y males were cultured in the presence of IL-2, IFNγ, IL-4, IL-6 and IL-12 for 4 days. The data represent 2 independent experiments.



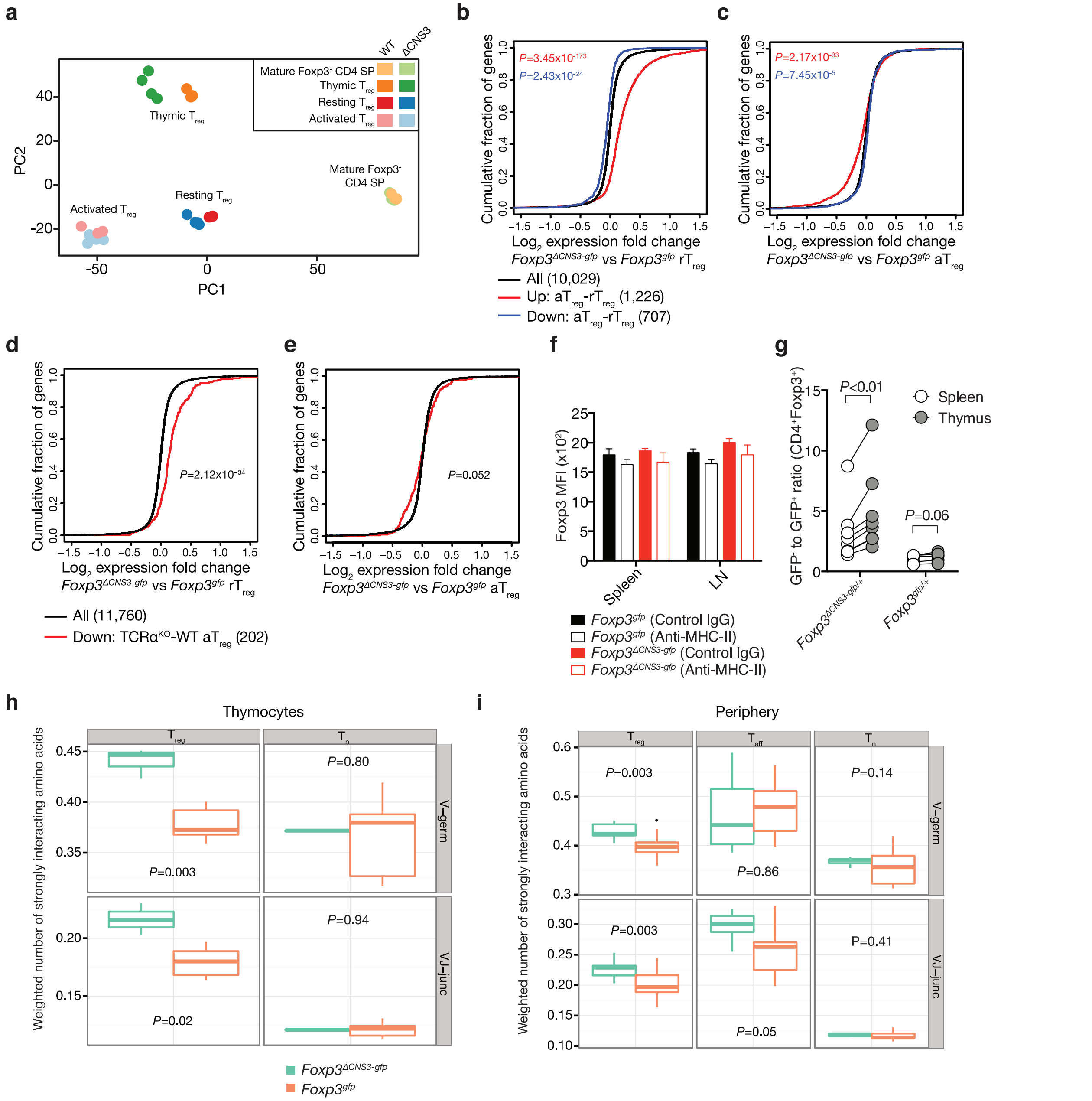
Extended Data Figure 2 | CNS3 is dispensable for the suppressor function of differentiated T_{reg} cells *in vivo*. **a**–**f**, *In vivo* assessment of the suppressor function of T_{reg} cells upon acute ablation of CNS3. T_{reg} cells ($CD4^+GFP^+$) isolated from *Foxp3*^{gfp} or *Foxp3*$^{CNS3\text{-}fl\text{-}gfp}$ mice were activated with CD3 and CD28 antibody-coated beads *in vitro* for five days and then transduced with retroviruses expressing Cre recombinase and a Thy1.1 reporter. Three days later, $Thy1.1^+CD4^+GFP^+$ cells were sorted by FACS for the suppressor assay. **a**, $CD4^+Foxp3^-$ and $CD8^+$ effector T cells (T_{eff}) sorted from *Foxp3*^{DTR} reporter mice seven days after diphtheria toxin (DT) injection (1 µg intraperitoneal per mouse) were transferred alone or with equal amounts of $Thy1.1^+$ Cre-transduced *Foxp3*^{gfp} or *Foxp3*$^{CNS3\text{-}fl\text{-}gfp}$ T_{reg} cells into *Tcrb*$^{-/-}$ *Tcrd*$^{-/-}$ recipients. **b**, Mice were weighed before and after T-cell transfer, and relative weight changes were assessed at weeks 3 and 4 post-transfer. **c**–**f**, Four weeks after adoptive transfer, cells were recovered and analysed for T_{reg} frequencies and Foxp3 expression (**c**), $CD4^+TCR\beta^+Foxp3^-$ and $CD8^+TCR\beta^+$ cell numbers (**d**), IFNγ (**e**) and IL-13 (**f**) production. Unpaired Mann–Whitney test ($n = 5$ per group).



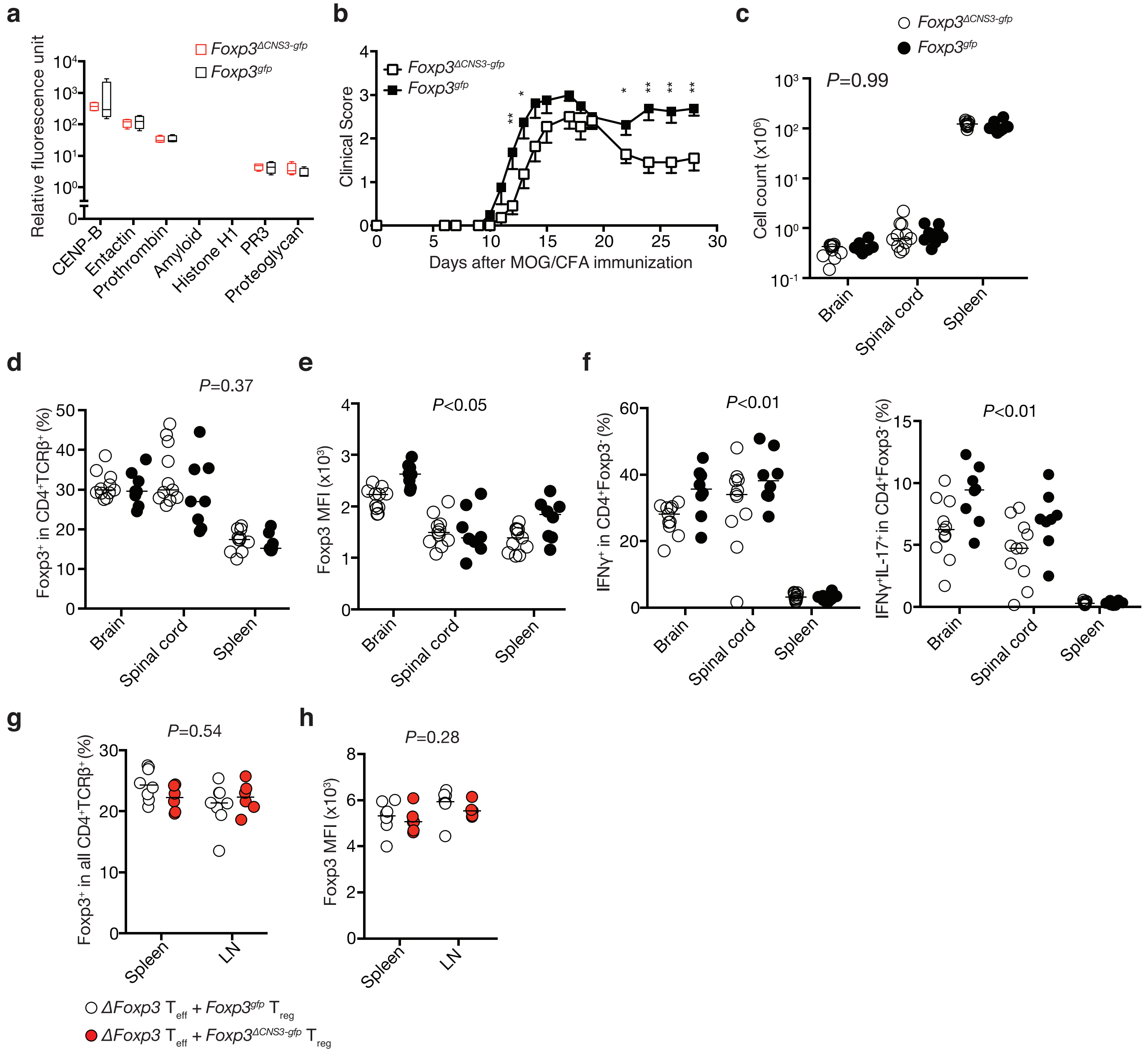
Extended Data Figure 3 | Epigenetic modifications at the *Foxp3* locus during T_{reg} differentiation. a, **b**, ChIP–qPCR analysis of H3K4me3 (**a**) and H3K27ac (**b**) at the *Foxp3* locus and control loci (*Hspa2*, *Rpl30* and *Gm5069*) in B cells, DP thymocytes, naive $CD4^+$ T and T_{reg} cells. FACS-sorted cells from wild-type male $Foxp3^{DTR}$ mice were used for ChIP–qPCR. Relative enrichment was calculated by normalizing to background binding to control region (*Gm5069*). **c**–**e**, ChIP–qPCR analysis of H3K4me1 (**c**), H3K4me3 (**d**) and H3K27ac (**e**) in the *Foxp3* locus in mature T_{reg} cells isolated from wild-type $Foxp3^{gfp}$ and $Foxp3^{\Delta CNS3\text{-}gfp}$ male mice normalized to the background binding to the *Gm5069* locus. **f**, CNS3-dependent deposition of H3K27ac at the *Foxp3* promoter in $Foxp3^-CD4^+$ T cells during *in vitro* T_{reg} cell induction. $Foxp3^{gfp}$ or $Foxp3^{\Delta CNS3\text{-}gfp}$ naive $CD4^+$ T cells were cultured under *in vitro* T_{reg} cell differentiation conditions. After three days of culture, GFP^- and GFP^+ cells were sorted for ChIP–qPCR analysis. Two-tailed unpaired *t*-test. **g**, Inhibition of T_{reg} induction *in vitro* by bromodomain protein inhibitor iBET. Naive $CD4^+$ T cells from wild-type $Foxp3^{gfp}$ males were used for Foxp3 *in vitro* induction in the presence of indicated concentrations of iBET or vehicle. **h**, Schematic of the chromatin dynamics at CNS3 and the *Foxp3* promoter during T_{reg} cell differentiation. The data are shown as means ± s.e.m. of triplicates and represent 1 of 2 independent experiments.



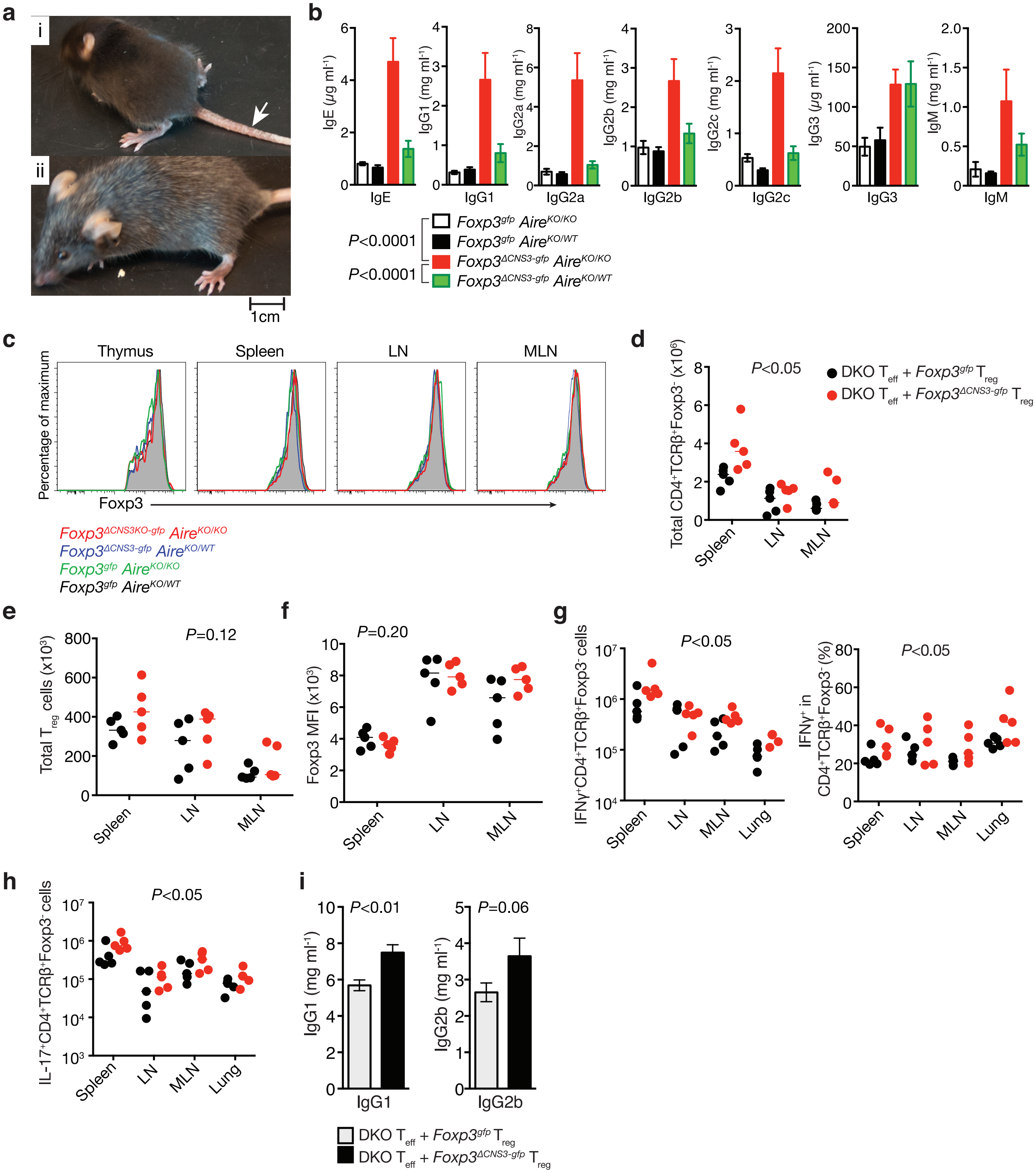
Extended Data Figure 4 | CNS3 facilitates Foxp3 induction and shapes T_{reg} cell repertoire. a, Differential effect of CNS3 on T_{reg} cell *in vitro* development of mature non-T_{reg} CD4 SP T cells. CD4 SP thymocytes ($CD4^+CD8^-TCR\beta^{hi}GFP^-CD25^-CD62L^{hi}CD69^{lo}$) were pooled and sorted from male *Foxp3*^{gfp} and *Foxp3*$^{\Delta CNS3\text{-}gfp}$ littermates ($n = 7$ each group) for *in vitro* T_{reg} cell induction performed with titrated CD3 antibody and lethally irradiated antigen-presenting cells isolated from wild-type B6 spleens in the presence of TGFβ and recombinant IL-2. Foxp3 expression was analysed four days later and the relative changes in the ratios of Foxp3-expressing cells in the absence of CNS3 were calculated by comparing to CNS3-sufficient groups. Data depict means ± s.e.m. of five replicate cultures and represent 1 of 3 independent experiments. **b**, Flow cytometric analysis of Nur77 protein expression in CNS3-deficient and -sufficient T_{reg} cells ($n = 5$ for each group). Two-tailed unpaired Mann–Whitney test. The data represent 1 of >2 independent experiments. **c**, Increased Nur77 protein levels in CNS3-deficient T_{reg} cells developed after conditional ablation of CNS3 upon tamoxifen-induced activation of *Ubc*$^{Cre\text{-}ERT2}$. Bone marrow of $CD45.1^+$ *Foxp3*^{gfp} and $CD45.2^+$ *Ubc*$^{Cre\text{-}ERT2}$ *Foxp3*$^{CNS3\text{-}fl\text{-}gfp}$ R26Y mice were collected from donor mice treated with tamoxifen, mixed at a 1:1 ratio and transferred into lethally irradiated *Tcrb*$^{-/-}$ *Tcrd*$^{-/-}$ recipients. $CD45.1^+CD4^+GFP^+$, $CD45.2^+YFP^-GFP^+$ and $CD45.2^+YFP^+GFP^+$ cells were sorted for flow cytometric analysis of Nur77 protein levels 10 weeks after bone marrow transfer ($n = 5$). Unpaired Mann–Whitney tests were used to compare $CD45.2^+YFP^+GFP^+$ and $CD45.2^+YFP^-GFP^+$ or $CD45.2^+YFP^+GFP^+$ and control ($CD45.1^+CD4^+GFP^+$) groups. The data show medians of individual mice and represent >3 independent experiments. **d**, Nur77 expression levels in thymic T_{reg} precursors ($CD25^+Foxp3^-$), immature ($CD62L^{lo}CD69^{hi}$) and mature ($CD62L^{hi}CD69^{lo}$) CD4 SP thymocytes, and peripheral $Foxp3^-CD4^+$ and $CD8^+$ T cells in 6–7-week-old *Foxp3*^{gfp} ($n = 5$) and *Foxp3*$^{\Delta CNS3\text{-}gfp}$ ($n = 4$) littermates. Unpaired Mann–Whitney test. The data show medians of individual mice and represent >3 independent experiments. **e**, Differential Nur77 expression in peripheral resting ($CD44^{lo}CD62L^{hi}$) and activated ($CD44^{hi}CD62L^{lo}$) T_{reg} cells (wild-type *Foxp3*^{gfp}). The data represent 1 of >3 independent experiments. **f**, **g**, Upregulation of Nur77 expression in resting ($CD44^{lo}CD62L^{hi}$) (**f**) and activated ($CD44^{hi}CD62L^{lo}$) (**g**) CNS3-deficient T_{reg} cells in 6–7-week-old *Foxp3*^{gfp} ($n = 5$) and *Foxp3*$^{\Delta CNS3\text{-}gfp}$ ($n = 4$) littermates. Unpaired Mann–Whitney test. The data represent >3 experiments. **h**, **i**, CTLA4 (**h**) and Ki-67 (**i**) expression by CNS3-deficient and -sufficient T_{reg} cells in *Foxp3*^{gfp} ($n = 9$) and *Foxp3*$^{\Delta CNS3\text{-}gfp}$ ($n = 11$) mice (**h**). *Foxp3*^{gfp} ($n = 5$); *Foxp3*$^{\Delta CNS3\text{-}gfp}$ ($n = 4$) (**i**). Two-tailed unpaired Mann–Whitney test. The data represent 1 of >3 independent experiments.



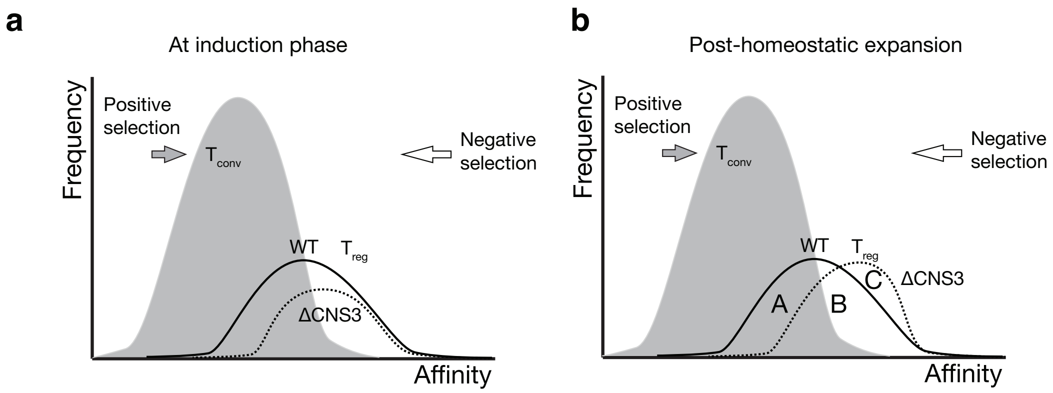
Extended Data Figure 5 | Influence of CNS3 on T_{reg} cell repertoire. **a**, Principal component analysis of mRNA expression in CNS3-deficient and -sufficient mature $Foxp3^-$ and $Foxp3^+$ CD4 SP thymocytes, and peripheral resting and activated T_{reg} cells. RNA-seq was performed with three and four biological replicates for cells sorted from male *Foxp3*^{gfp} and *Foxp3*$^{\Delta CNS3\text{-}gfp}$ littermates, respectively. Dots represent samples from individual mice. **b**, **c**, Relative gene expression levels (cumulative fraction of genes) in CNS3-sufficient and -deficient peripheral resting (**b**) or activated (**c**) T_{reg} cells in comparison to those up- and downregulated in activated versus resting T_{reg} cells isolated from *Foxp3*^{gfp} mice. The numbers of genes in each comparison group are indicated in parentheses. **d**, **e**, Relative gene expression levels in CNS3-sufficient and -deficient peripheral rT_{reg} (**d**) or aT_{reg} (**e**) cells in comparison to those downregulated in activated T_{reg} cells subjected to acute TCR ablation versus mock treatment. The numbers of genes in each comparison group are indicated in parentheses. One-tailed Kolmogorov–Smirnov test. **f**, Flow cytometric analysis of Foxp3 expression level (median fluorescence intensity (MFI)) in CNS3-sufficient and -deficient T_{reg} cells after expansion in lymphopenic recipients. T_{reg} cells were sorted from mixed bone marrow chimaeras of $CD45.1^+$ *Foxp3*^{gfp} and $CD45.2^+$ *Foxp3*$^{\Delta CNS3\text{-}gfp}$ mice and mixed at a 1:1 ratio, and co-transferred with wild-type naive $Foxp3^-CD4^+$ T cells into *Tcrb*$^{-/-}$ *Tcrd*$^{-/-}$ recipients treated with MHC-II-blocking antibody or isotype-control IgG before and after the transfer ($n = 5$ per group). Mean ± s.e.m; the data represent 1 of 3 independent experiments. Unpaired *t*-test revealed no statistically significant difference between matched CNS3-deficient and -sufficient groups ($P > 0.3$). **g**, Comparison of CNS3-sufficient and -deficient T_{reg} cells in competitive environment of heterozygous *Foxp3*$^{gfp/+}$ and *Foxp3*$^{\Delta CNS3\text{-}gfp/+}$ females (6–8 weeks of age). In contrast to CNS3-sufficient T_{reg} cells, CNS3-deficient cells are relatively enriched in the periphery in comparison with the thymus. Ratios of GFP^- to GFP^+ T_{reg} cells are inversely proportional to the relative abundance of *Foxp3*^{gfp} or *Foxp3*$^{\Delta CNS3\text{-}gfp}$ T_{reg} cells in the T_{reg} pool. Wilcoxon matched-pairs signed rank test; *Foxp3*$^{gfp/+}$ ($n = 5$), *Foxp3*$^{\Delta CNS3\text{-}gfp/+}$ ($n = 8$). Linked circles represent samples from the same mice. Data represent 1 of 2 independent experiments. **h**, **i**, Numbers of strongly interacting amino acid residues (LFIMVWCY) were calculated for the V-segment of TCRα CDR3 (binned to germline) and V–J segment junction, and weighted by the corresponding clonotype frequencies. Sums of the weighted scores were used for the comparisons between CNS3-deficient and -sufficient groups (unpaired *t*-test). The data represent the analysis of pooled TCR sequences derived from the indicated thymic (**h**) and peripheral (**i**) $CD4^+$ naive (T_n), activated effector (T_{eff}) and T_{reg} cell subsets isolated from individual *Foxp3*^{gfp} ($n = 5$) and *Foxp3*$^{\Delta CNS3\text{-}gfp}$ ($n = 3$) mice. Box-and-whisker plots show minimum, maximum, first and third quartiles and median.



Extended Data Figure 6 | Selective modulation of autoimmune responses in mice lacking CNS3. **a**, CNS3 deficiency does not affect antibody production against a subset of autoantigens. *Foxp3*$^{\Delta CNS3\text{-}gfp}$ and *Foxp3*^{gfp} littermates ($n = 4$ per group). Box-and-whisker plots show minimum, maximum, first and third quartiles and median. Data represent 1 of 2 independent experiments. **b**–**f**, CNS3 deficiency decreases experimental autoimmune encephalomyelitis severity. On immunization with MOG peptide in CFA, mice of indicated genotypes were assessed for the severity of limb paralysis (**b**), effector T-cell numbers (**c**), T_{reg} cell frequency (**d**), Foxp3 expression levels (**e**) and inflammatory cytokine production (**f**). *Foxp3*^{gfp} ($n = 8$); *Foxp3*$^{\Delta CNS3\text{-}gfp}$ ($n = 11$). Unpaired *t*-test (**b**) or Mann–Whitney test (**c**–**f**). Mean and s.e.m. are presented (**b**). $*P < 0.05$, $**P < 0.01$. The data represent 2 independent experiments. **g**, **h**, Analysis of the proportion of T_{reg} cells in $CD4^{+}TCR\beta^{+}$ cell population (**g**) and level of Foxp3 expression (MFI) (**h**) in an *in vivo* suppressor assay of CNS3-deficient or -sufficient T_{reg} cells (Fig. 3e–h). Two-tailed unpaired Mann–Whitney test.



Extended Data Figure 7 | Compromised suppressive function of CNS3-deficient T_{reg} cells. **a**, Autoimmune diseases in *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/KO}$ (DKO) mice. Arrow indicates the inflammatory lesions in the tail of a 3-week-old mouse with an early onset of autoimmunity ($n > 11$) (i). De-pigmentation in a 6-week-old mouse with delayed onset of autoimmunity ($n > 16$) (ii). **b**, Analysis of serum Ig isotypes in *Foxp3*$^{\Delta CNS3\text{-}gfp}$ *Aire*$^{KO/KO}$ and littermate control mice using ELISA ($n = 8$ per group). Error bars, mean ± s.e.m. Two-way ANOVA. **c**, Flow cytometric analysis of Foxp3 expression by T_{reg} cells. The data show one of at least three mice per group and represent >3 independent experiments. **d**–**i**, Analysis of the ability of CNS3-deficient and -sufficient T_{reg} cells to control CNS3 *Aire* DKO effector T cells on adoptive transfer into T-cell-deficient recipients. Flow cytometric analysis of non-T_{reg} $CD4^+$ T-cell numbers (**d**), T_{reg} cell numbers (**e**), Foxp3 expression levels (**f**), IFNγ production (**g**), IL-17 production (**h**), and serum IgG1 and IgG2b levels (**i**) in recipient mice transferred with CNS3 and *Aire* DKO effector T cells ($Foxp3^-CD4^+$ and $CD8^+$) at a 10:1 ratio with T_{reg} cells from *Aire*-sufficient *Foxp3*^{gfp} or *Foxp3*$^{\Delta CNS3\text{-}gfp}$ mice. Two-tailed unpaired Mann–Whitney tests (**d**–**h**) or unpaired *t*-test (**i**). Error bars, mean ± s.e.m. (**i**). The recipient mice were analysed 7 weeks after adoptive T-cell transfer ($n = 5$ per group).



Extended Data Figure 8 | Theoretical impact of CNS3 on Treg TCR repertoire . a, Hypothetical distribution of TCRs expressed by T_{reg} and non-T_{reg} $CD4^+$ T cells according to their affinities for self-antigens. Precursor cells expressing TCRs within a certain low-affinity window are positively selected and become 'conventional' $CD4^+$ T cells, and those with higher affinities for self-antigens differentiate into T_{reg} cells. CNS3 promotes the differentiation of T_{reg} cells and broadens their TCR repertoire by facilitating Foxp3 expression predominantly in response to lower strength ('suboptimal') inducing TCR signals. **b**, After expansion in the periphery, CNS3-deficient T_{reg} cells reach similar numbers as their wild-type counterparts, with some TCRs underrepresented (A), some minimally affected (B), and some overrepresented (C). T_{conv}, conventional T cells.

doi:10.1038/nature16151

Depletion of fat-resident T_{reg} cells prevents age-associated insulin resistance

Sagar P. Bapat[1,2], Jae Myoung Suh[2,3], Sungsoon Fang[2,4], Sihao Liu[2], Yang Zhang[1], Albert Cheng[1], Carmen Zhou[1], Yuqiong Liang[1], Mathias LeBlanc[2], Christopher Liddle[5], Annette R. Atkins[2], Ruth T. Yu[2], Michael Downes[2], Ronald M. Evans[2,6] & Ye Zheng[1]

Age-associated insulin resistance (IR) and obesity-associated IR are two physiologically distinct forms of adult-onset diabetes. While macrophage-driven inflammation is a core driver of obesity-associated IR[1–6], the underlying mechanisms of the obesity-independent yet highly prevalent age-associated IR[7] are largely unexplored. Here we show, using comparative adipo-immune profiling in mice, that fat-resident regulatory T cells, termed fT_{reg} cells, accumulate in adipose tissue as a function of age, but not obesity. Supporting the existence of two distinct mechanisms underlying IR, mice deficient in fT_{reg} cells are protected against age-associated IR, yet remain susceptible to obesity-associated IR and metabolic disease. By contrast, selective depletion of fT_{reg} cells via anti-ST2 antibody treatment increases adipose tissue insulin sensitivity. These findings establish that distinct immune cell populations within adipose tissue underlie ageing- and obesity-associated IR, and implicate fT_{reg} cells as adipo-immune drivers and potential therapeutic targets in the treatment of age-associated IR.

The young, lean state is associated with insulin sensitivity, while both ageing and obesity can lead to the development of IR (Extended Data Fig. 1a). To explore key immune cell types that drive age- versus obesity-associated IR, we quantitatively profiled the immune cell components of adipose depots using a flow cytometry approach termed adipo-immune profiling (AIP) (Extended Data Fig. 1b–d and Extended Data Table 1). In contrast to the decrease in anti-inflammatory M2 adipose tissue macrophages and eosinophils observed in obesity-driven IR, AIP revealed that these cell populations are largely unperturbed in visceral adipose tissue (VAT) from aged mice[8–12] (M2 adipose tissue macrophages, aged: 33.6 ± 3.8% (mean ± s.d.), young: 29.8 ± 4.1%, obese: 22.9 ± 6.3%; eosinophils, aged: 4.4% ± 1.6%, young: 4.7% ± 0.7%, obese: 0.8% ± 1.0%; Fig. 1a). Instead, the relative portion of the non-macrophage compartment is significantly increased in aged compared to young or obese mice (aged: 24.3 ± 4.6%, young: 17.9 ± 2.8%, obese: 15.7 ± 3.8%; Fig. 1a), which is largely attributable to an ~12-fold expansion in the fT_{reg} cell population[13,14] (aged: 5.0 ± 1.2%, young: 0.4 ± 0.1%, obese: 0.1 ± 0.1%; Fig. 1a, b). These condition-dependent AIP signatures of adipose tissue suggest that distinct pathophysiologic processes drive age- and obesity-associated IR and specifically implicate fT_{reg} cells in age-associated IR.

T_{reg} cells in fat express *Pparg* at a high level, which allows them to expand their relative numbers approximately 6–7-fold[15]. Knockout of *Pparg* in T_{reg} cells blocks this accumulation. Accordingly, we exploited this observation by creating *Foxp3*^{Cre} (*Foxp3*-IRES-YFP-Cre) *Pparg*$^{fl/fl}$ mice in which T_{reg} cells are selectively depleted (from 6.1% to 0.9%) from VAT[15] (fT_{reg} knockout mice; Fig. 2a and Extended Data Fig. 2a, b), although the depletion of PPARγ-positive T_{reg} cells in tissues such as muscle and liver cannot be ruled out. This depletion is achieved without significantly altering the immune profiles of subcutaneous adipose tissue (SAT) or spleen (Extended Data Fig. 2c, d). Importantly, the 6.8-fold VAT-specific loss of fT_{reg} cells does not elicit any overt signs of systemic inflammation generally associated with T_{reg} cell dysfunction. Aged fT_{reg} knockout mice have normal-sized spleens and increased $CD62L^{hi}$ $CD44^{lo}$ naive $CD4^+$ T-cell populations compared to wild-type controls (Fig. 2c and Extended Data Fig. 3a). The normal intestinal histology provides additional evidence that the T_{reg} cell population is not perturbed (Extended Data Fig. 3b, c). Furthermore, no differences are observed in the levels of inflammatory cytokines, including TNFα, IL-1β, IL-6, IFNγ and IL-17, in the serum of aged fT_{reg} knockout compared to control mice (Extended Data Fig. 3d).

Notably, the selective loss of fT_{reg} cells attenuates many of the hallmarks of age-associated metabolic dysregulation[16]. They weigh less

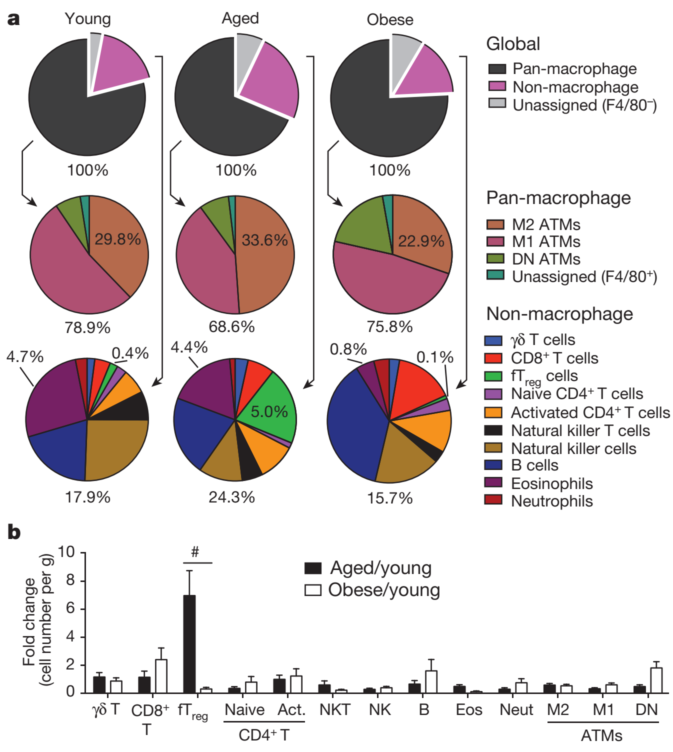


Figure 1 | fT_{reg} cells are selectively enriched in aged mice. a, VAT AIPs from mice at 12 weeks (young, $n = 10$), 44 weeks (aged, $n = 10$) and in diet-induced obese mice ($n = 10$). Immune cell abundance is expressed as percentage of $CD45.2^+$ cells. ATM, adipose tissue macrophages; DN, double negative. **b,** Changes in immune cell abundance between indicated groups, expressed as fold change in cell number per gram of VAT. Obese mice were fed a HFD for 12 weeks from 12 weeks of age. NK, natural killer cells; NKT, natural killer T cells. Data are mean ± s.e.m. #, false discovery rate <2%.

[1]Immunobiology and Microbial Pathogenesis Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA. [2]Gene Expression Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA. [3]Graduate School of Medical Science and Engineering, KAIST 34141, South Korea. [4]Department of Biotechnology, College of Life Sciences, Sejong University, Seoul 143-747, South Korea. [5]Storr Liver Centre, Westmead Millennium Institute, Sydney Medical School, University of Sydney, Sydney 2145, Australia. [6]Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, California 92037, USA.

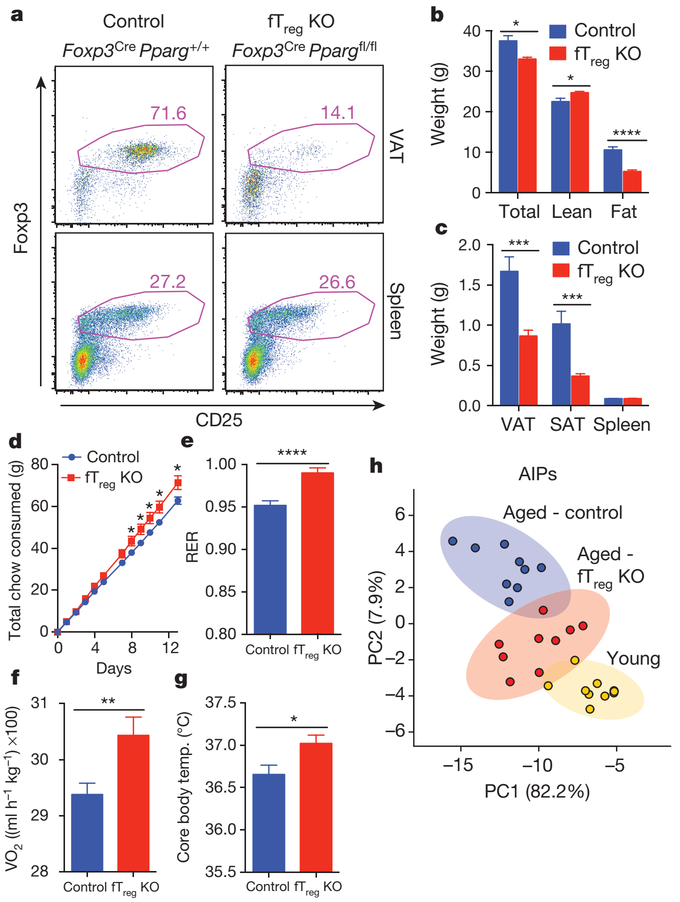


Figure 2 | fT_{reg} knockout mice are protected from general hallmarks of metabolic ageing. a, Representative FACS plots of fT_{reg} knockout (KO) (*Foxp3*^{Cre} *Pparg*$^{fl/fl}$) and control (*Foxp3*^{Cre} *Pparg*$^{+/+}$) mice depicting T_{reg} cell enrichment in VAT and spleen (~15 months, CD45.2$^+$ CD4$^+$ gating). **b**, Total body weight ($n = 15$ per group), and lean and fat mass of control and fT_{reg} knockout mice (~12 months, $n = 8$ per group). **c**, Mass of VAT, SAT and spleen in aged control and fT_{reg} knockout mice (~15 months, $n = 9$ per group). **d**, Cumulative food consumption of control and fT_{reg} knockout mice (~8–9 months, $n = 8$ per group). **e**, **f**, Average 24 h respiratory exchange ratio (RER) (**e**) and average oxygen consumption (VO_2) (**f**) of aged control and fT_{reg} knockout mice (~11 months, $n = 6$ per group). **g**, Core body temperature of control and fT_{reg} mice (~13 months, $n = 9$ per group). **h**, Principal component analysis of non-macrophage AIPs of young (12 weeks), aged (~15 months) and aged fT_{reg} knockout (~15 months) mice ($n = 9$ per group). Data are mean ± s.e.m. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, Student's t-test.

than control mice and are leaner (decreased VAT and SAT adiposity) despite increased food consumption (Fig. 2b–d). In addition, the respiratory exchange ratio (Fig. 2e), oxygen consumption (Fig. 2f) and core body temperature (Fig. 2g) are increased in aged fT_{reg} knockout mice compared to control mice. These marked improvements suggest that the age-associated metabolic phenotype is closely linked with VAT immune responses, and that in the aged setting, a reduction in fT_{reg} cell numbers may be protective. Indeed, the AIPs of aged fT_{reg} knockout mice are shifted towards those of young mice, as visualized by principal component analysis (Fig. 2h).

Although the fT_{reg} knockout phenotype is most pronounced in aged mice, a reduction in fT_{reg} cell levels can also be found in obese fT_{reg} knockout mice (Fig. 3a). However, the beneficial metabolic effects of fT_{reg} ablation are only significant in age-associated metabolic dysregulation, in which fasting serum glucose and insulin levels are significantly reduced (Fig. 3b, c and Extended Data Fig. 4a). Furthermore, aged fT_{reg} knockout mice display smaller glucose excursions during glucose tolerance tests and increased sensitivity during insulin tolerance tests compared to weight-matched control mice (Fig. 3e). Again, these improvements in glucose homeostasis are observed only in aged mice; no significant differences are seen in young or obese fT_{reg} knockout mice (Fig. 3d, f), which is consistent with the largely unchanged AIPs of obese fT_{reg} knockout mice (Extended Data Fig. 5a, b).

Although fT_{reg} cells were previously implicated in the insulin-sensitizing function of the PPARγ agonists thiazolidinediones (TZDs)[15], in our studies fT_{reg} knockout mice display similar metabolic improvements to the TZD rosiglitazone as control mice. These beneficial effects of TZDs are evident with either direct treatment of obese mice (therapeutic intervention; Extended Data Fig. 6a–g) or prophylactic treatment (drug intervention coincident with high-fat diet (HFD) feeding; Extended Data Fig. 6h–l). Additionally, we find that the insulin-sensitizing effects of rosiglitazone precede the TZD-induced expansion of fT_{reg} cells in HFD-fed mice (Extended Data Fig. 6m–q). While intercolony variation cannot be ruled out, our findings do not support a significant role for fT_{reg} cells in the therapeutic mechanism of action of TZDs.

Histologically, aged fT_{reg} knockout VAT depots appear similar to control mice, and inflammatory processes such as macrophage crowning are observed at comparable frequencies (Fig. 3g and data not shown). However, aged fT_{reg} knockout VAT has increased levels of TNFα (Extended Data Fig. 7a), increased expression of *Vegfa* (implicated in adipose remodelling and insulin sensitivity[17]; Extended Data Fig. 7b) and decreased expression of extracellular matrix genes (including collagen VI implicated in adipose tissue rigidity[18], and the wound response gene *Sparc*; Extended Data Fig. 7b, c) compared to control tissue. Accompanying these changes, several proteases involved in extracellular matrix remodelling and angiogenesis (members of the ADAM, ADAMTS, MMP and CELA families) are differentially expressed (Extended Data Fig. 7b, d). Of note, adipocytes from aged fT_{reg} knockout mice are smaller than those in control mice (fT_{reg} knockout: ~70% $< 5{,}000\,\mu m^2$, control: ~41% $< 5{,}000\,\mu m^2$; Fig. 3h and Extended Data Fig. 4b), and serum non-esterified free fatty acid levels are reduced to almost half those of control mice; both indicators of improved insulin sensitivity (Fig. 3i). In addition, circulating levels of the adipokine resistin, which positively correlates with mouse IR, are reduced in the aged fT_{reg} knockout mice[19,20] (Fig. 3j). Furthermore, aged fT_{reg} knockout mice present with decreased hepatic steatosis, as determined histologically and by decreased fasting hepatic and serum triglyceride content (Fig. 3l–n). In combination, these findings suggest that the loss of fT_{reg} cells in adipose tissue alleviates many of the indications of age-associated IR in mice, a primary clinical manifestation of metabolic ageing.

To associate fT_{reg} cells more directly with age-associated IR, we measured basal glucose uptake in adipose tissue *ex vivo*. Notably, VAT from fT_{reg} knockout mice took up almost twice the amount of glucose as control tissue (Fig. 3k). Conversely, expansion of fT_{reg} cells in wild-type mice via treatment with IL-2–IL-2-monoclonal-antibody complex[21] abrogates basal glucose uptake in VAT by ~50% (Fig. 3o, p). This inverse correlation between fT_{reg} cell numbers and glucose uptake in adipose tissue supports a causal association between fT_{reg} cells and IR during ageing.

Our findings of an association between fT_{reg} cells and age-associated IR and metabolic ageing suggest that these cells are functionally distinct from splenic T_{reg} cells. To investigate this notion, we compared the transcriptomes of T_{reg} cells, as well as conventional CD4$^+$ T (T_{conv}) cells, isolated from VAT and spleen. Comparative analyses revealed that while certain canonical genes are similarly expressed (for example, *Foxp3*, *Ctla4* and *Tigit*), VAT and splenic T_{reg} cells have discrete expression signatures, consistent with the suggested functional distinction. In particular, *Pparg*, *Gata3* and *Irf4* are selectively enriched in VAT but not splenic T_{reg} cells[22] (Extended Data Fig. 8a). Furthermore, unbiased comparative gene expression analyses combined with hierarchical clustering defined extensive fat- and splenic-residence clusters (1,142 and 1,431 genes, respectively) relative to

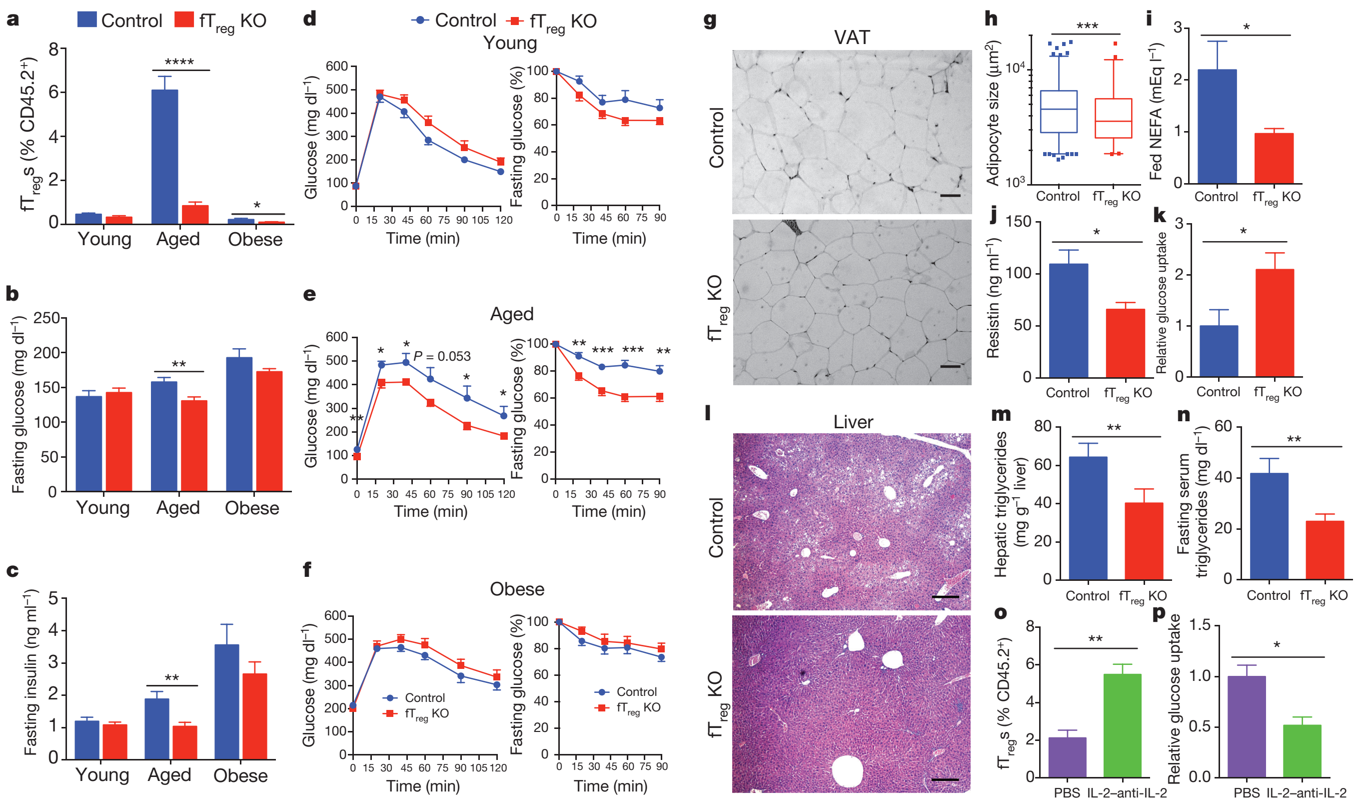


Figure 3 | Loss of fT$_{reg}$ cells protects against the clinical hallmarks of age-associated IR. a, fT$_{reg}$ levels in VAT from control and fT$_{reg}$ knockout (*Foxp3*^{Cre} *Pparg*$^{fl/fl}$) mice in young (12 weeks, control $n=6$; fT$_{reg}$ KO $n=6$), aged (~15 months, control $n=10$; fT$_{reg}$ KO $n=10$) and obese (control $n=6$; fT$_{reg}$ KO $n=8$) cohorts. **b**, Fasting serum glucose (**b**) and insulin levels (**c**) in control and fT$_{reg}$ knockout mice in young (12 weeks, control $n=9$; fT$_{reg}$ KO $n=9$), aged (36 weeks, control $n=9$; fT$_{reg}$ KO $n=11$) and obese (control $n=10$; fT$_{reg}$ KO $n=10$) cohorts. **d–f**, Glucose (left) and insulin (right) tolerance tests of control and fT$_{reg}$ knockout mice in young (12 weeks, control $n=8$; fT$_{reg}$ KO $n=8$) (**d**), aged (36–37 weeks, control $n=8$; fT$_{reg}$ KO $n=9$) (**e**) and obese (control $n=9$; fT$_{reg}$ KO $n=10$) (**f**) cohorts. **g**, Representative haematoxylin and eosin (H&E) staining of VAT (epididymal) from ~14-month-old control ($n=3$) and fT$_{reg}$ knockout mice ($n=5$). Scale bars, 50 μm. **h**, Box and whisker plot of adipocyte size distribution in VAT from control ($n=3$) and fT$_{reg}$ knockout mice ($n=3$) (~14 months). **i**, Ad-libitum-fed serum non-esterified fatty acid (NEFA) levels in ~14-month-old control ($n=9$) and fT$_{reg}$ knockout mice ($n=10$). **j**, Serum resistin levels in ~14-month fasted control and fT$_{reg}$ knockout mice ($n=4$ pooled samples (2 mice per sample) per group). **k**, Post-prandial glucose uptake in VAT of aged control ($n=5$) and fT$_{reg}$ knockout mice ($n=4$). **l**, Representative H&E staining of liver from ~14-month-old control ($n=3$) and fT$_{reg}$ knockout mice ($n=5$). Scale bars, 200 μm. **m**, Hepatic triglyceride levels in ~14-month-old control ($n=5$) and fT$_{reg}$ knockout mice ($n=3$). **n**, Fasting serum triglycerides in ~14-month-old control ($n=9$) and fT$_{reg}$ knockout mice ($n=10$). **o**, fT$_{reg}$ cells, expressed as percentage of total CD45.2^{+} cells, in control (PBS) and IL-2–anti-IL-2-treated mice ($n=3$ mice per group). **p**, Relative glucose uptake in VAT of 16-week-old control and IL-2–anti-IL-2-treated mice ($n=4$ mice per group). Data are mean ± s.e.m. $*P<0.05$, $**P<0.01$, $***P<0.001$, $****P<0.0001$, Student's *t*-test.

much smaller pan-T$_{reg}$ clusters 1 and 2 (56 and 162 genes, respectively). Transcriptionally, fT$_{reg}$ cells cluster more closely with fat T$_{conv}$ cells than splenic T$_{reg}$ cells (Fig. 4a), suggesting that the functional specification of fT$_{reg}$ cells is informed by their anatomical location within adipose tissue, as well as the expression of the T$_{reg}$ cell-lineage-specifying transcription factor Foxp3 (refs 23, 24 and Fig. 4b). Importantly, aged fT$_{reg}$ cells maintain their suppressive functionality as measured by *in vitro* suppression assays (Fig. 4c, d), and indicated by the high expression levels of *Ctla4* (ref. 25), *Il2ra* (ref. 26), and the anti-inflammatory cytokine *Il10* (Fig. 4b). We posit that the transcriptional differences between fT$_{reg}$ cells and splenic T$_{reg}$ cells (found in the fT$_{reg}$ cell cluster of 1,049 genes) may provide a therapeutic route to manipulate fT$_{reg}$ cell populations selectively. The IL-33 receptor ST2, which lies within the fT$_{reg}$ cell cluster, has been recently implicated in effector T$_{reg}$ and in particular fT$_{reg}$ cell development[27,28]. Indeed, ST2 was ~60 and ~30 times more highly expressed in fT$_{reg}$ cells than in splenic T$_{reg}$ and fat T$_{conv}$ cells, respectively, consistent with the ImmGen database (http://www.immgen.org; Fig. 4e and Extended Data Fig. 8b). Flow cytometry confirmed that ST2 is expressed on the cell surface of most fT$_{reg}$ cells, but on relatively few fat T$_{conv}$ cells, or splenic T$_{reg}$ or T$_{conv}$ cells (Fig. 4f, g). Furthermore, VAT has ~25 times more ST2^{+} fT$_{reg}$ cells than ST2^{+} fat T$_{conv}$ cells; a similarly trending ~10 times difference is observed in the spleen (Fig. 4h).

To explore the therapeutic potential of the IL-33/ST2 signalling pathway, aged mice were initially injected with IL-33 (0.5 μg intraperitoneally (i.p.) on days 0, 2 and 4, analysis on day 6) to expand the fT$_{reg}$ cell population (Fig. 4i–k). In agreement with fT$_{reg}$ cell expansion driven by IL-2–anti-IL-2 treatment, mice injected with IL-33 display signs of IR (basal glucose uptake in VAT reduced to ~60% of control mice; Fig. 4l). In the converse approach, acute treatment with an anti-ST2 antibody (200 μg per mouse i.p. on days 0 and 2, analysis on day 3) is able to significantly deplete fT$_{reg}$ cells (~50% reduction), with a smaller percentage reduction of splenic T$_{reg}$ cells (Fig. 4m). Notably, the partial depletion of fT$_{reg}$ cells achieved with acute anti-ST2 treatment coincides with an increase in insulin-stimulated glucose uptake in VAT (~25% increase in glucose uptake compared to control treated mice; Fig. 4n), suggesting a link between fT$_{reg}$ cell depletion and increased adipose insulin sensitivity. Furthermore, this increase in insulin sensitivity is achieved without any signs of T$_{conv}$ cell activation associated with systemic T$_{reg}$ cell dysfunction (Extended Data Fig. 8c–e).

As obesity and ageing often associate in humans, we challenged aged fT$_{reg}$ knockout and control mice with HFD. While initially protected against HFD-induced weight gain and associated metabolic dysregulation, the metabolic benefits attributed to the loss of fT$_{reg}$ cells were progressively lost over 8 weeks (Extended Data Fig. 9a–e), further

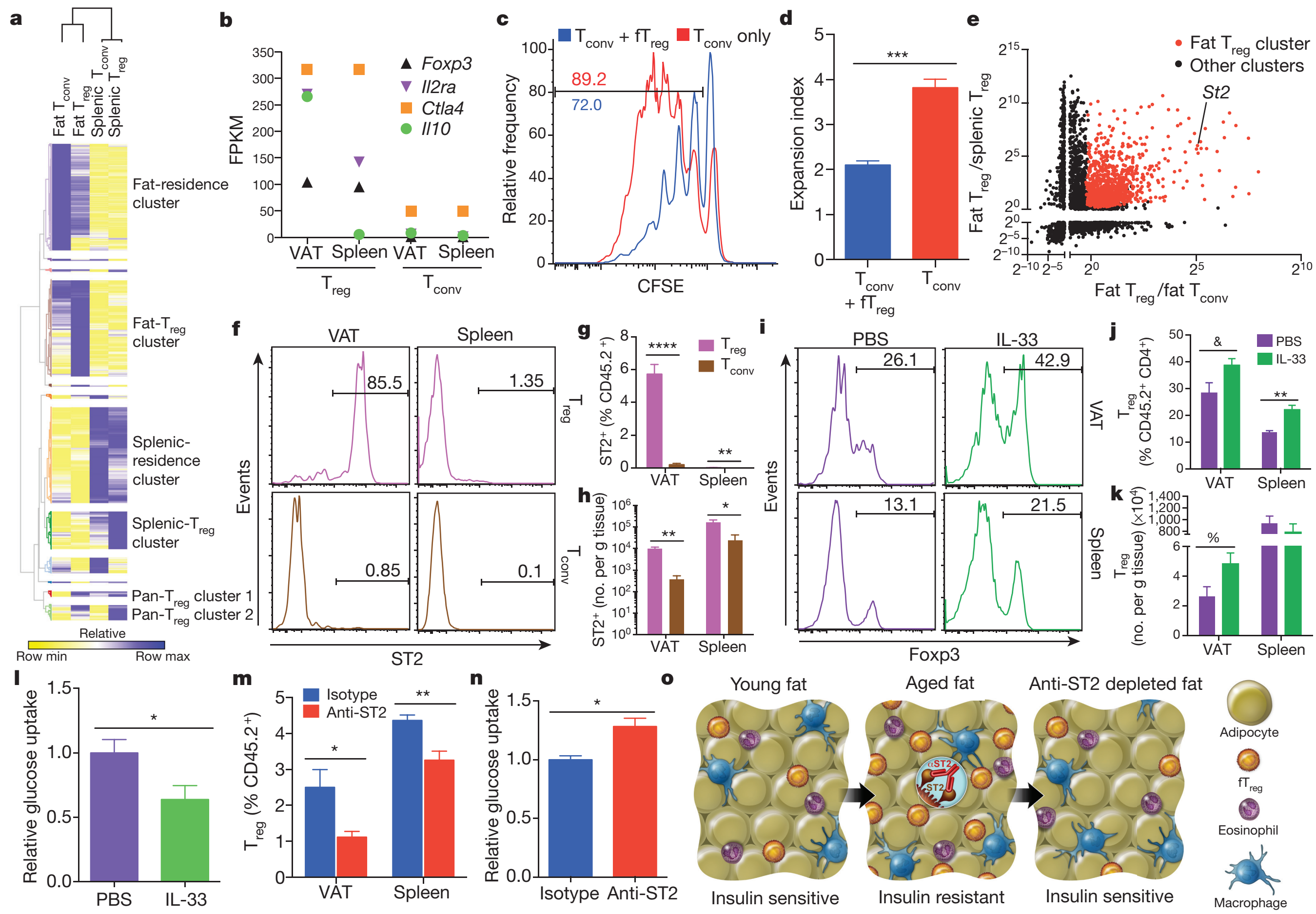


Figure 4 | fT_{reg} cell depletion improves adipose glucose uptake. **a**, Hierarchical clustering of differentially expressed genes between fat T_{reg} and T_{conv} and splenic T_{reg} and T_{conv} cells from *Foxp3*$^{Thy1.1}$ mice (47 weeks, cells pooled from 3–4 mice, same data set used in **b**, **e**). **b**, Fragments per kilobase of transcripts per million mapped reads (FPKM) values of selected genes important for T_{reg} cell identity and canonical suppressive function. **c**, **d**, *In vitro* suppression assay of fT_{reg} cells (pooled from retired breeders, added at 1:1 ratio with splenic T_{conv} cells, conducted in triplicate). **c**, Representative carboxyfluorescein succinimidyl ester (CFSE) tracings of T_{conv} cells with or without fT_{reg} cells. Gating indicates percentage of dividing cells. **d**, Expansion index of T_{conv} cells. **e**, Fold change in expression levels of differentially expressed genes across fat T_{reg} and T_{conv} and splenic T_{reg} and T_{conv} cells. Fat T_{reg} cell cluster genes are labelled in red. Position of *St2* (also known as *Il1rl1*) is marked. **f**, **g**, Representative FACS analysis (**f**) and quantification of ST2 expression (**g**) in $CD4^+$ T cells from aged mice (45 weeks, $n = 5$ mice). **h**, Total number of $ST2^+$ T_{reg} and T_{conv} cells per gram of tissue in VAT and spleen ($n = 5$ mice). **i**–**k**, FACS histograms (**i**) and quantification (**j**) of T_{reg} cells (percentage $Foxp3^+$ of $CD45^+$ $CD4^+$ population), and cells per gram of tissue in VAT and spleen after IL-33 or PBS treatment (**k**) (16 weeks, $n = 5$ mice per group). **l**, *Ex vivo* glucose uptake in VAT from wild-type mice after control or IL-33 treatment (16 weeks, $n = 5$ mice per group). **m**, Quantification of fT_{reg} and splenic T_{reg} cells (percentage $Foxp3^+$ of $CD45^+$ population) (**m**) and *ex vivo* insulin-stimulated glucose uptake (**n**) in VAT from wild-type mice after anti-ST2 depleting antibody or isotype control treatment (~45 weeks, $n = 4$ mice per group). **o**, Adipo-immune model of metabolic ageing. Data are mean ± s.e.m. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$, &$P = 0.053$, %$P = 0.056$, Student's *t*-test.

suggesting that distinct pathophysiologies drive obesity- versus age-associated IR.

Taken together, our data provide evidence that distinct adipo-immune populations orchestrate unique features of age- and obesity-associated IR. We show in the non-obese setting that *Pparg*-positive fT_{reg} cells accumulate to unusually high levels (6.7%) as a function of age, exacerbating both the decline of adipose metabolic function as well as the rise in IR (Fig. 4o). These results are in marked contrast to the increased role of M1 adipose tissue macrophages in metabolic dysfunction linked to obesity, coupled with a suppression of fT_{reg} levels to 0.9%. Thus, these studies highlight contrasting roles of the immune compartment in contributing to key aspects of adipose health and disease.

Given the classical immune suppressive and anti-inflammatory nature of T_{reg} cells, we speculate that the chronic inflammatory processes that drive obesity-associated IR seem unlikely to be driving age-associated IR. Indeed, there is increasing appreciation that maintaining a certain degree of inflammation is beneficial for adipose tissue remodelling and its metabolic function[29]. Failure to preserve an optimal immune state in the aged adipose tissue may directly contribute to metabolic disorders such as IR and age-associated diabetes. We suggest type IV diabetes as a designation for non-obese-dependent fT_{reg}-driven metabolic disease of the elderly.

In this context, it is of particular significance that fT_{reg} cells in aged adipose tissue express the cytokine receptor ST2 at ~30–60-fold higher levels than in other sites such as spleen, making the fT_{reg} cell population sensitive to depletion via anti-ST2 treatment. While ST2 has been implicated in other physiological processes and immune cell types that may also affect glucose homeostasis, this simple fT_{reg} cell depletion approach increases adipose tissue insulin sensitivity suggesting the potential of selective fT_{reg} cell depletion therapy in the prevention of age-related IR.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 May; accepted 28 October 2015.
Published online 18 November 2015.

1. Ferrante, A. W. Jr. Macrophages, fat, and the emergence of immunometabolism. *J. Clin. Invest.* **123,** 4992–4993 (2013).
2. Lumeng, C. N. & Saltiel, A. R. Inflammatory links between obesity and metabolic disease. *J. Clin. Invest.* **121,** 2111–2117 (2011).
3. Mathis, D. Immunological goings-on in visceral adipose tissue. *Cell Metab.* **17,** 851–859 (2013).
4. Osborn, O. & Olefsky, J. M. The cellular and signaling networkslinking the immune system and metabolism in disease. *Nature Med.* **18,** 363–374 (2012).
5. Weisberg, S. P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest.* **112,** 1796–1808 (2003).
6. Xu, H. Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J. Clin. Invest.* **112,** 1821–1830 (2003).
7. CDC. 2014 National Diabetes Statistics Report; http://www.cdc.gov/diabetes/data/statistics/2014statisticsreport.html (2014).
8. Wu, D. *et al.* Eosinophils sustain adipose alternatively activated macrophages associated with glucose homeostasis. *Science* **332,** 243–247 (2011).
9. Qiu, Y. *et al.* Eosinophils and type 2 cytokine signaling in macrophages orchestrate development of functional beige fat. *Cell* **157,** 1292–1308 (2014).
10. Rao, R. R. *et al.* Meteorin-like is a hormone that regulates immune-adipose interactions to increase beige fat thermogenesis. *Cell* **157,** 1279–1291 (2014).
11. Odegaard, J. I. *et al.* Macrophage-specific PPAR-γ controls alternative activation and improves insulin resistance. *Nature* **447,** 1116–1120 (2007).
12. Fujisaka, S. *et al.* Regulatory mechanisms for adipose tissue m1 and m2 macrophages in diet-induced obese mice. *Diabetes* **58,** 2574–2582 (2009).
13. Feuerer, M. *et al.* Lean, but not obese, fat is enriched for a unique population of regulatory T cells that affect metabolic parameters. *Nature Med.* **15,** 930–939 (2009).
14. Lumeng, C. N. *et al.* Aging is associated with an increase in T cells and inflammatory macrophages in visceral adipose tissue. *J. Immunol.* **187,** 6208–6216 (2011).
15. Cipolletta, D. *et al.* PPAR-γ is a major driver of the accumulation and phenotype of adipose tissue T_{reg} cells. *Nature* **486,** 549–553 (2012).
16. Houtkooper, R. H. *et al.* The metabolic footprint of aging in mice. *Sci. Rep.* **1,** 134 (2011).
17. Sun, K. *et al.* Dichotomous effects of VEGF-A on adipose tissue dysfunction. *Proc. Natl Acad. Sci. USA* **109,** 5874–5879 (2012).
18. Khan, T. *et al.* Metabolic dysregulation and adipose tissue fibrosis: role of collagen VI. *Mol. Cell. Biol.* **29,** 1575–1591 (2009).
19. Steppan, C. M. *et al.* The hormone resistin links obesity to diabetes. *Nature* **409,** 307–312 (2001).
20. Schwartz, D. R. & Lazar, M. A. Human resistin: found in translation from mouse to man. *Trends Endocrinol. Metab.* **22,** 259–265 (2011).
21. Webster, K. E. *et al.* *In vivo* expansion of T reg cells with IL-2-mAb complexes: induction of resistance to EAE and long-term acceptance of islet allografts without immunosuppression. *J. Exp. Med.* **206,** 751–760 (2009).
22. Feuerer, M. *et al.* Genomic definition of multiple *ex vivo* regulatory T cell subphenotypes. *Proc. Natl Acad. Sci. USA* **107,** 5919–5924 (2010).
23. Josefowicz, S. Z., Lu, L.-F. & Rudensky, A. Y. Regulatory T cells: mechanisms of differentiation and function. *Annu. Rev. Immunol.* **30,** 531–564 (2012).
24. Ohkura, N., Kitagawa, Y. & Sakaguchi, S. Development and maintenance of regulatory T cells. *Immunity* **38,** 414–423 (2013).
25. Wing, K. *et al.* CTLA-4 control over Foxp3^{+} regulatory T cell function. *Science* **322,** 271–275 (2008).
26. Fontenot, J. D. *et al.* A function for interleukin 2 in Foxp3-expressing regulatory T cells. *Nature Immunol.* **6,** 1142–1151 (2005).
27. Vasanthakumar, A. *et al.* The transcriptional regulators IRF4, BATF and IL-33 orchestrate development and maintenance of adipose tissue–resident regulatory T cells. *Nature Immunol.* **16,** 276–285 (2015).
28. Schiering, C. *et al.* The alarmin IL-33 promotes regulatory T-cell function in the intestine. *Nature* **513,** 564–568 (2014).
29. Asterholm, I. W. *et al.* Adipocyte inflammation is essential for healthy adipose tissue expansion and remodeling. *Cell Metab.* **20,** 103–118 (2014).

Acknowledgements We would like to thank L. Chong, J. Alvarez, Y. Dai, S. Kaufman and B. Collins for technical assistance, L. Ong and C. Brondos for administrative assistance, and J. Simon for assistance with graphics. S.P.B. is supported by National Institutes of Health (NIH) grants F30 DK096828 and T32 GM007198. C.L. and M.D. are funded by grants from the National Health and Medical Research Council of Australia Project grants 512354, 632886 and 1043199. R.M.E. is an Investigator of the Howard Hughes Medical Institute (HHMI) at the Salk Institute and March of Dimes Chair in Molecular and Developmental Biology, and is supported by NIH grants DK057978, DK090962, HL088093, HL105278 and ES010337, the Glenn Foundation for Medical Research, the Leona M. and Harry B. Helmsley Charitable Trust, Ipsen/Biomeasure, California Institute for Regenerative Medicine and The Ellison Medical Foundation. Y. Zheng is supported by the Nomis Foundation, the Rita Allen Foundation, the Emerald Foundation, the Hearst Foundation, the National Multiple Sclerosis Society, and National Institutes of Health (AI099295 and AI107027). This work was also supported by National Cancer Institute funded Salk Institute Cancer Center core facilities (CA014195) and the James B. Pendleton Charitable Trust.

Author Contributions S.P.B., J.M.S., M.D., R.M.E. and Y. Zheng designed and supervised the research. S.P.B., J.M.S., S.F., S.L., Y. Zhang, A.C., C.Z. and Y.L. performed research. S.P.B., J.M.S., M.L., C.L., A.R.A., R.T.Y., M.D., R.M.E. and Y. Zheng analysed data. S.P.B., J.M.S., S.L., A.R.A., R.T.Y., M.D., R.M.E. and Y. Zheng wrote the manuscript.

Author Information RNA-Seq data can be accessed in the NCBI Sequence Read Archive under the accession SRP053799. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.M.E. (evans@salk.edu) or Y.Z. (yzheng@salk.edu).

METHODS

Mice. All mice were housed in the specific pathogen-free facilities at The Salk Institute for Biological Studies or purchased from Taconic Biosciences. C57BL/6NTac mice were purchased from Taconic Biosciences for comparative AIPs or studies that required wild-type aged mice. Age-matched retired breeders were purchased for AIPs of aged adipose tissue, and diet-induced obese C57BL/6NTac mice were purchased for profiling of obese adipose tissue. fT_{reg} knockout mice were generated by crossing *Foxp3*^{Cre} (*Foxp3*-IRES-YFP-Cre)[30] and *Pparg*$^{fl/fl}$ (ref. 31) mice. We used the *Foxp3*$^{Thy1.1}$ (ref. 32) reporter mice when isolating T_{reg} and T_{conv} $CD4^+$ cells from spleen and fat for subsequent RNA-Seq analysis. Mice within The Salk Institute for Biological Studies received autoclaved normal chow (MI laboratory rodent diet 5001, Harlan Teklad), irradiated HFD (60 kcal% fat, Research Diets), or irradiated HFD with rosiglitazone (30 mg kg^{-1} of food, Research Diets). All mice used for studies were male. All procedures involving animals were performed in accordance with protocols approved by the Institutional Animal Care and Use Committee (IACUC) and Animal Resources Department (ARD) of the Salk Institute for Biological Studies.

AIPs. Visceral (epididymal) and subcutaneous (inguinal) adipose depots were dissected from mice after 10 ml PBS perfusion through the left ventricle. Inguinal lymph nodes resident in inguinal adipose tissue were removed. Adipose tissue was minced into fine pieces (2–5 mm^3) and digested in stromal vascular isolation buffer (100 mM HEPES, pH 7.4, 120 mM NaCl, 50 mM KCl, 5 mM glucose, 1 mM $CaCl_2$ and 1.5% BSA) containing 1 mg ml^{-1} collagenase at 37 °C with intermittent shaking for 1.5 h. The suspension was then passed through a 100-μm mesh to remove undigested clumps and debris. The flow-through was allowed to stand for 10 min to separate the floating adipocyte fraction and infranatant containing the stromal vascular fraction. The infranatant was removed while minimally disturbing the floating adipocyte fraction and centrifuged at 400*g* for 10 min. The pellet containing the stromal vascular fraction was washed once in 10 ml RPMI. The resultant isolated cells were subjected to FACS analysis. The following antibodies were used to assemble the adipo-immune profile. BioLegend: CD45.2 (104), CD44 (IM7), CD62L (MEL-14), TCRg/d (GL3), CD19 (6D5), CD25 (PC61), CD206 (C068C2), CD301 (LOM-14); eBioscience: CD3 (145-2C11), CD25 (PC61), CD4 (RM4–5), TCRb (H57–597), B220 (RA3-6B2), NK1.1 (PK136), CD49b (DX5), Foxp3 (FJK-16 s), F4/80 (BM8), CD11c (N418), CD11b (M1/70); Tonbo biosciences: F4/80 (BM8.1), CD4 (RM4-5), CD44 (IM7), CD62L (MEL-14), Ly6G (RB6-8C5); BD Pharmingen: Siglec-F (E50–2440); BD Biosciences: CD8a (53-6.7). Cells were analysed using the BD FACSAria instrument (Becton Dickinson) and FlowJo software (Tree Star).

Body composition and adipocyte size analyses. Body composition was measured with an Echo MRI-100 body composition analyser (Echo Medical Systems). VAT (epididymal) was dissected, and the wet weight was determined. Adipose tissues were fixed in 10% formalin, sectioned, and stained in haematoxylin and eosin. An adipocyte cross-sectional area was determined from photomicrographs of VAT using ImageJ.

***In vivo* metabolic phenotype analysis.** Real-time metabolic analyses were conducted in a Comprehensive Lab Animal Monitoring System (Columbus Instruments). CO_2 production, O_2 consumption and ambulatory counts were determined for at least three consecutive days and nights after at least 24 h for adaptation before data recording.

Principal component analysis of AIPs. Non-macrophage immune cell populations, described as percentage of the total $CD45.2^+$ immune compartment, were inputted into MetaboAnalyst 3.0 (http://www.metaboanalyst.ca/MetaboAnalyst/) for PCA. No normalizations, transformations, or scalings were implemented.

Glucose homeostasis studies. Fasting was induced for 6 h, except for glucose tolerance tests, which were conducted after overnight fasting. Glucose (1–2 g kg^{-1}, i.p.) and insulin (0.5–1.0 U kg^{-1}, i.p.) was injected for glucose tolerance tests and insulin tolerance tests, respectively. Blood glucose was monitored using a Nova Max Plus glucometer.

Histological analyses. Sections (4 μm) of fixed tissues were stained with haematoxylin and eosin according to standard procedures. Histopathological scores were graded on blinded samples for severity and extent of inflammation and morphological changes by a pathologist.

Serum analyses. Blood was collected by tail bleeding or right atrial puncture. Non-esterified fatty acids (Wako) and triglycerides (Thermo) were measured using colorimetric methods. Serum insulin levels (Ultra Sensitive Insulin, Crystal Chem) were measured by ELISAs. Serum cytokine and metabolic hormone levels were analysed by the Luminex Bio-Plex system using the Mouse Cytokine 23-Plex Panel and Diabetes Panel, respectively, according to the manufacturer's instructions (Bio-Rad).

Core body temperature. Mice were single housed, and core body temperature was measured with a clinical rectal thermometer (Thermalert model TH-5; Physitemp) during the middle of the light cycle. The probe was dipped in a room temperature lubricating glycerol before insertion.

***Ex vivo* 2-DG uptake assays.** Adipose tissue was dissected from mouse, cut into small pieces with scissors, washed and incubated for 30 min with Krebs-Ringer Bicarbonate HEPES buffer (KRBH, 120 mM NaCl, 4 mM KH_2PO_4, 1 mM $MgSO_4$, 0.75 mM $CaCl_2$, 30 mM HEPES and 10 mM $NaHCO_3$, pH 7.4, supplemented with 1% fatty-acid-free BSA). For determination of exogenous insulin-stimulated 2-deoxy-D-glucose (2-DG) uptake, adipose was incubated in KRBH with 100–200 nM insulin for 20 min at 37 °C. Cold 2-DG and hot 2-DG-1,2-^{3}H(N) was added to incubated adipose tissue such that the final concentration of cold 2-DG was 0.1 mM and final quantity of hot 2-DG-1,2-^{3}H(N) was 0.1 μCi (assuming total reaction volume ~400 μl). Adipose was further incubated for 20 min at 37 °C, then washed three times with PBS before being lysed by scintillation fluid. 2-DG uptake was determined by measuring scintillation counts normalized to adipose tissue mass used for assay. Nonspecific 2-DG uptake levels were determined by co-treating adipose tissue with cytochalasin B (0.1 μM final concentration) with the addition of cold and hot 2-DG.

IL-2–anti-IL-2 complex and IL-33 injections. IL-2–anti-IL-2 complexes were prepared by incubating 2 μg of mouse IL-2 (Biolegend) with 10 μg of anti-IL-2 antibody (JES6.1, Bioxcell) in a total volume of 200 μl PBS for 30 min at 37 °C (amounts given per injection). Mice were injected i.p. three times (days 0, 1 and 2) and analysed on day 8. For IL-33 expansion assays, mice were injected i.p. with 0.5 μg of recombinant mouse IL-33 in PBS (R&D systems) three times (days 0, 2 and 4) and analysed on day 6. PBS was used for control injections.

RNA-Seq library generation. Total RNA was isolated from sorted cells or whole tissues using TRIzol reagent (Invitrogen) as per the manufacturer's instructions and treated with DNaseI (Qiagen) for 30 min at 22 °C. Sequencing libraries were prepared from 10–100 ng of total RNA using the TruSeq RNA sample preparation kit v2 (Illumina) according to the manufacturer's protocol. In brief, mRNA was purified, fragmented and used for first- and second-strand cDNA synthesis followed by adenylation of 3′ ends. Samples were ligated to unique adaptors and subjected to PCR amplification. Libraries were then validated using the 2100 BioAnalyzer (Agilent), normalized and pooled for sequencing. RNA-Seq libraries prepared from two biological replicates for each experimental condition were sequenced on the Illumina HiSeq 2500 using barcoded multiplexing and a 100-bp read length.

High-throughput sequencing and analysis. Image analysis and base calling were done with Illumina CASAVA-1.8.2. This yielded a median of 29.9 M usable reads per sample. Short read sequences were mapped to a UCSC mm9 reference sequence using the RNA-Seq aligner STAR[33]. Known splice junctions from mm9 were supplied to the aligner and *de novo* junction discovery was also permitted. Differential gene expression analysis, statistical testing and annotation were performed using Cuffdiff 2 (ref. 34). Transcript expression was calculated as gene-level relative abundance in fragments per kilobase of exon model per million mapped fragments and employed correction for transcript abundance bias[35]. RNA-Seq results for genes of interest were also explored visually using the UCSC Genome Browser.

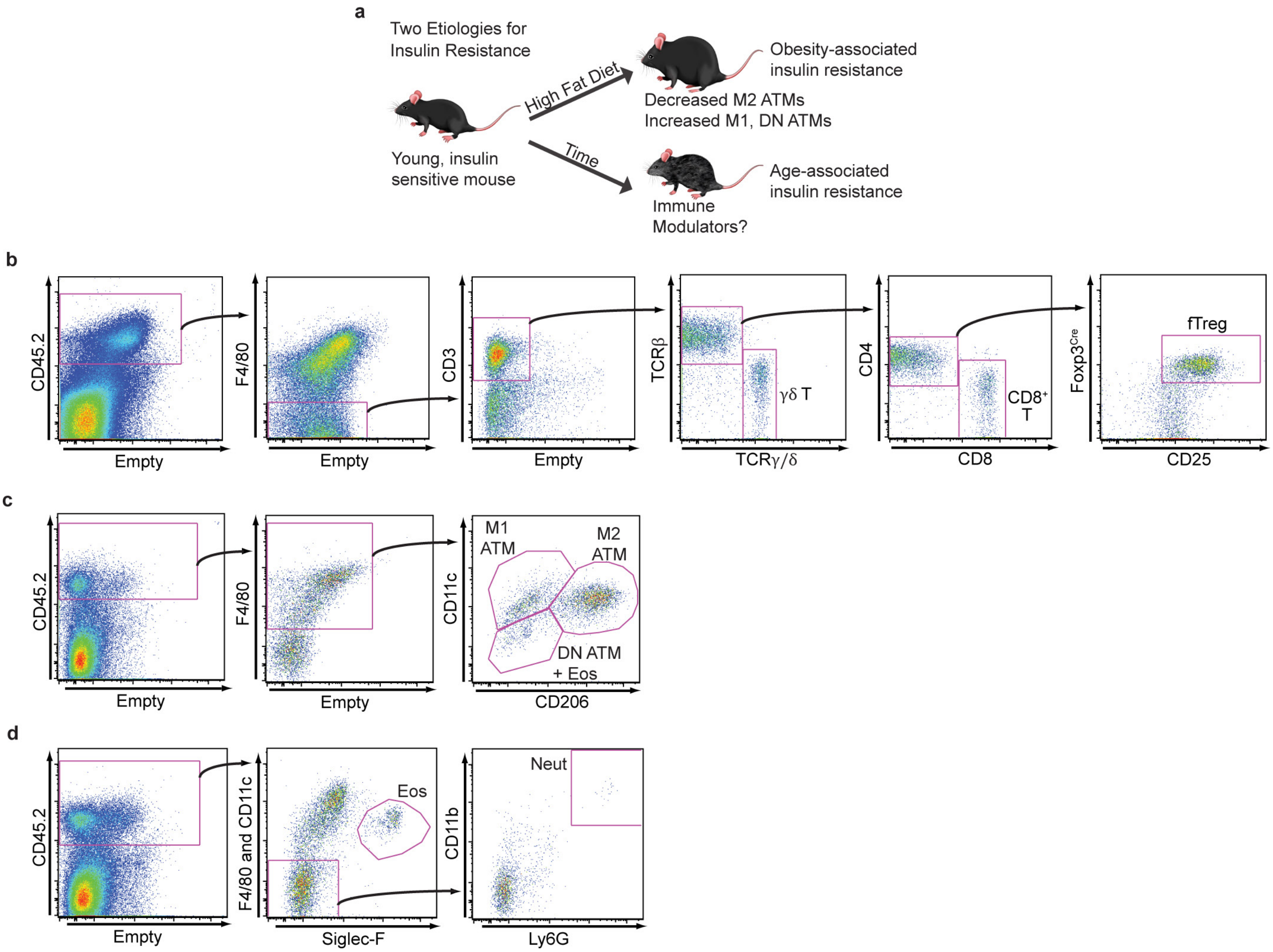
Hierarchical clustering. Differentially expressed gene names and corresponding FPKM values across samples were inputted into GENE-E (Broad Institute) for hierarchical clustering analysis (implemented one minus pearson correlation for sample and gene distance metrics and the average linkage method) and visualization. Gene cluster names were created to describe the gene expression characteristics within each cluster (that is, fat-residence cluster refers to the gene cluster in which genes are expressed at greater levels in T cells residing in fat. Fat-T_{reg} cluster refers to the gene cluster in which genes are expressed highest in only the fT_{reg} cells).

***In vitro* suppression assay.** fT_{reg} cells were isolated from aged wild-type mice treated with IL-2–anti-IL-2 complexes to expand fT_{reg} cell numbers, as described above with isolation conducted on day 6. Stromal vascular fractions were isolated from VAT as described above, and fT_{reg} cells were sorted from stromal vascular fractions using the BD FACSAria instrument (Becton Dickinson), gating on $CD45.2^+$ $CD4^+$ $CD25^+$ cells. $CD45.1^+$ mice were used to isolate splenic responder T cells, which were purified by positive selection using CD4-specific Dynabeads (Invitrogen), followed by sorting on a BD FACSAria cell sorter, gating on $CD45.1^+$ $CD4^+$ $CD25^-$ $CD62L^{hi}$ $CD44^{lo}$ cells. Antigen-presenting cells were prepared from wild-type B6 splenocytes by T-cell depletion using Thy1-specific MACS beads. CFSE labelled effector T cells (5×10^4 cells $well^{-1}$) were co-cultured with fT_{reg} cells at the indicated ratio in the presence of irradiated (30 Gy) antigen-presenting cells (1×10^5 cells $well^{-1}$) in 96-well plates in complete RPMI1640 medium supplemented with 10% FBS and CD3 antibody (1 μg ml^{-1}). Cell proliferation and expansion index were determined 96 h later using the BD FACSAria instrument and analysed with the FlowJo software package (Tree Star).

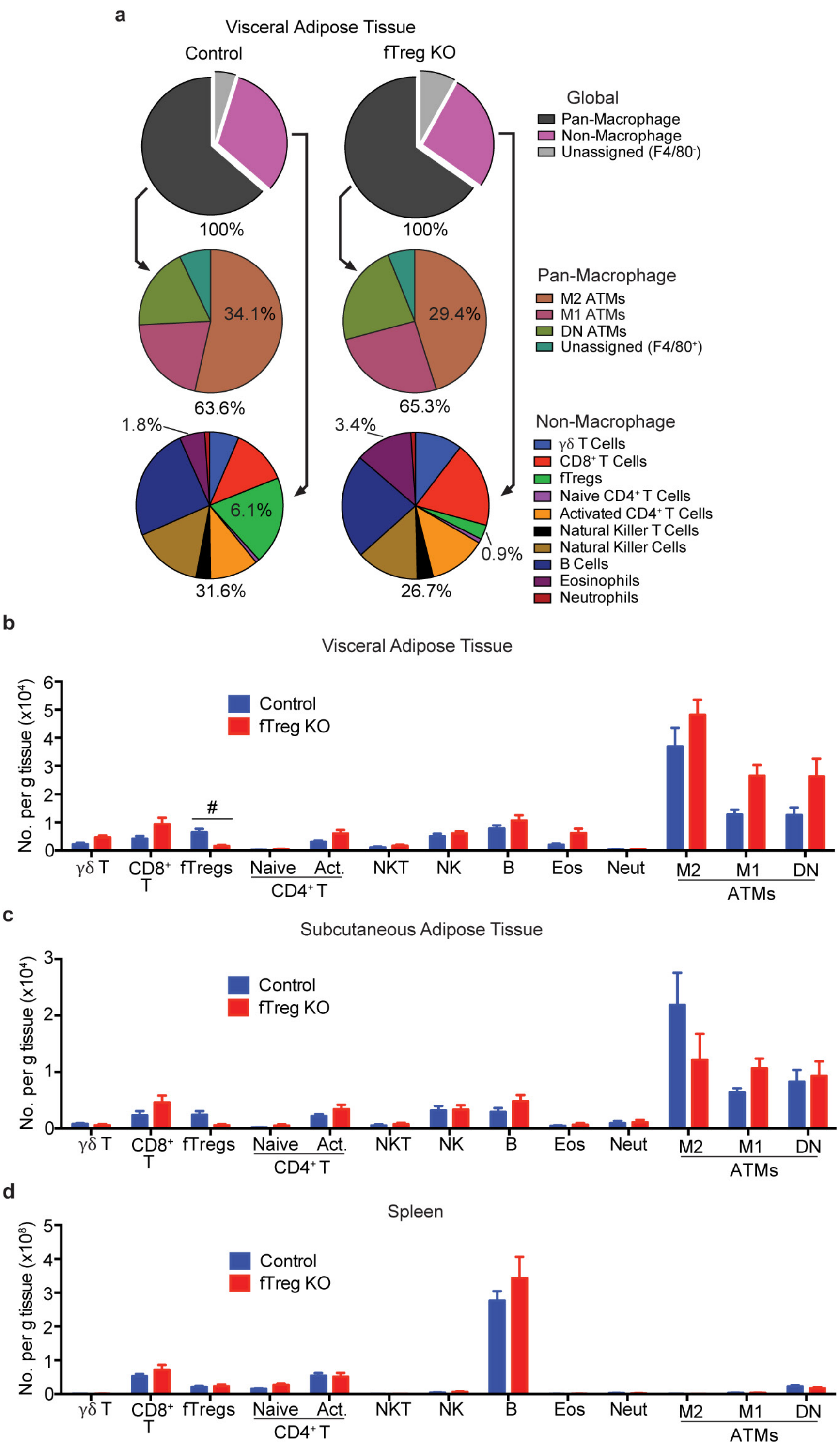
ST2 studies and anti-ST2-depleting antibody treatment. Antibody for ST2 FACS analysis was purchased from MD Bioproducts, clone DJ8. For fT_{reg} cell depletion, mice were injected i.p. with 200 μg anti-ST2 antibodies[36] (R&D systems, clone 245707) or isotype control (Bioxcell) twice (days 0 and 2) and euthanized for analysis on day 3.

Statistical analyses. Statistical analyses were performed with Prism 6.0 (GraphPad). P values were calculated using two-tailed unpaired or paired Student's t-test. When analysing AIPs, we used a false discovery rate approach to avoid the problem of an inflated false positive rate due to the substantial number of hypothesis tests. Mice cohort size was designed to be sufficient to enable statistical significance to be accurately determined. When applicable, mice were randomly assigned to treatment or control groups. No animals were excluded from the statistical analysis, with the exception of exclusions due to technical errors, and the investigators were not blinded in the studies. Appropriate statistical analyses were applied, assuming a normal sample distribution, as specified in the figure legends. No estimate of variance was made between each group. All *in vivo* metabolic and glucose homeostasis experiments, *ex vivo* glucose uptake experiments, and AIP experiments were conducted with at least two independent cohorts. RNA-Seq experiments, Luminex profiling and histological analyses were conducted using multiple biological samples (as indicated in figure legends) from indicated cohorts.

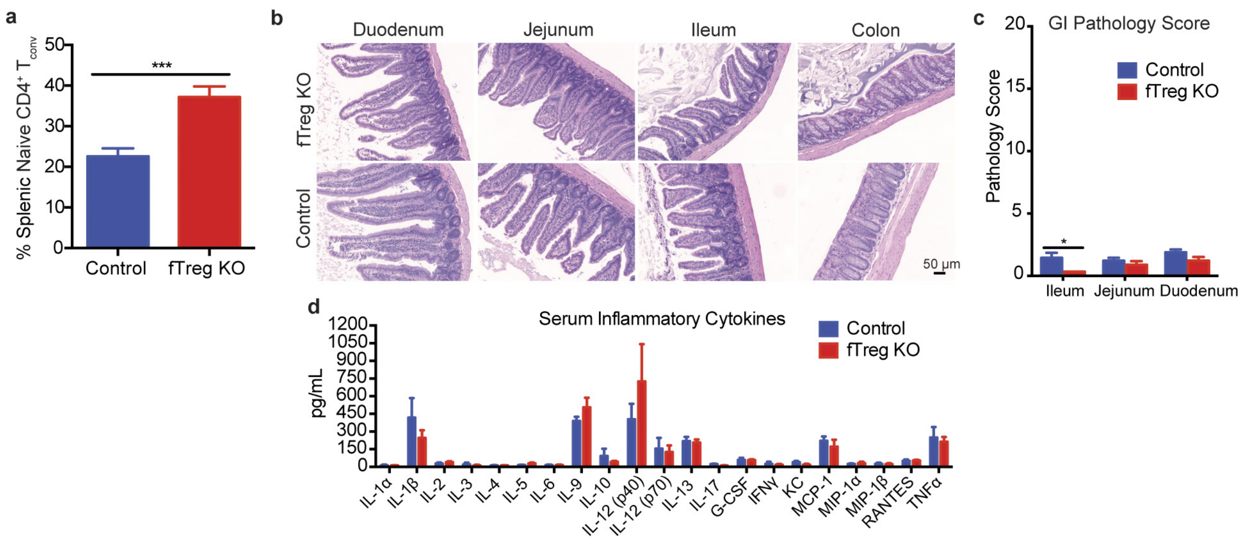
30. Rubtsov, Y. P. *et al.* Regulatory T cell-derived interleukin-10 limits inflammation at environmental interfaces. *Immunity* **28,** 546–558 (2008).
31. He, W. *et al.* Adipose-specific peroxisome proliferator-activated receptor gamma knockout causes insulin resistance in fat and liver but not in muscle. *Proc. Natl Acad. Sci. USA* **100,** 15712–15717 (2003).
32. Liston, A. *et al.* Differentiation of regulatory Foxp3$^+$ T cells in the thymic cortex. *Proc. Natl Acad. Sci. USA* **105,** 11903–11908 (2008).
33. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
34. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnol.* **31,** 46–53 (2012).
35. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27,** 2325–2329 (2011).
36. Monticelli, L. A. *et al.* Innate lymphoid cells promote lung-tissue homeostasis after infection with influenza virus. *Nature Immunol.* **12,** 1045–1054 (2011).



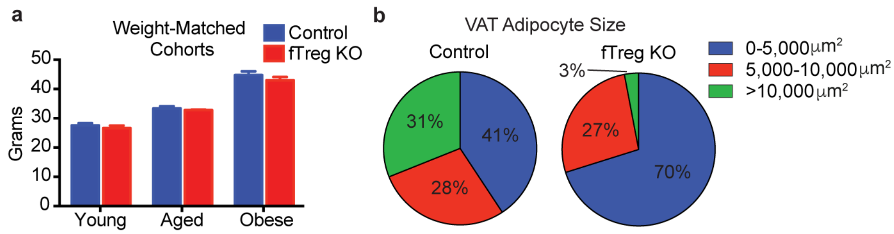
Extended Data Figure 1 | Schematic outlining study premise and selected gating strategies used to generate AIPs. a, Schematic outlining study premise. **b–d**, AIPs were generated through the use of several distinct antibody cocktails. Here, using *Foxp3*$^{\mathrm{Cre}}$ (*Foxp3*-IRES-YFP-Cre) reporter mice, we show how the stromal vascular fraction of VAT was analysed by flow cytometry to identify several T cell subtypes (**b**), macrophage subsets (**c**) and eosinophils and neutrophils (**d**).



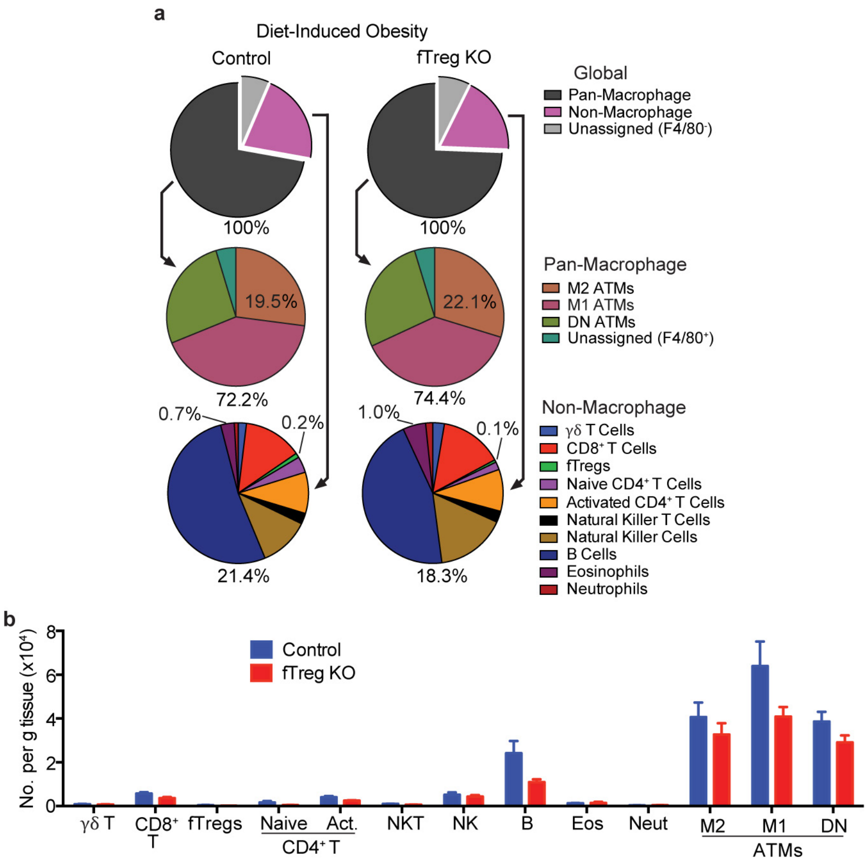
Extended Data Figure 2 | AIPs of aged fT$_{reg}$ knockout and control mice in VAT, SAT and spleen. a, AIPs of aged (~14 months) fT$_{reg}$ knockout and control male mice depicting immune cell abundance, expressed as percentage of CD45.2$^+$ cells. Entirety of immune compartment (top) is further divided into pan-macrophage (middle) and non-macrophage (bottom) pie charts ($n = 9$ mice per group). **b–d**, Immune cell abundance between fT$_{reg}$ knockout and control mice, expressed as cells per gram of VAT (**b**), SAT (**c**) and spleen (**d**) ($n = 9$ mice per group). Data are mean ± s.e.m. #, false discovery rate < 2%.



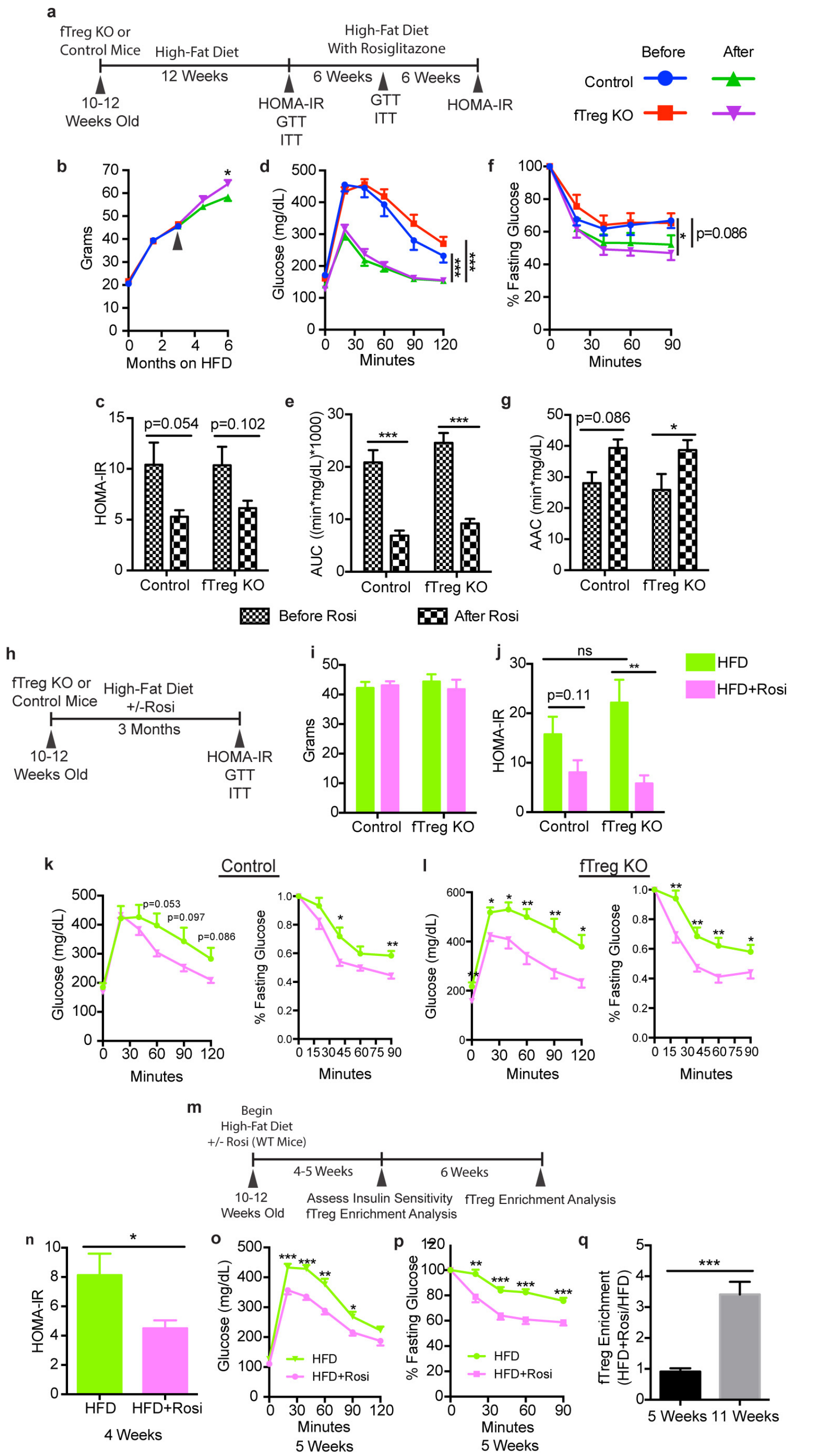
Extended Data Figure 3 | Aged fT$_{reg}$ knockout mice do not show signs of systemic autoimmunity or breakdown in peripheral tolerance. **a**, Percentage of splenic naive CD4$^+$ T cells as defined by CD62^{hi} CD44^{lo} relative to total CD4$^+$ CD25$^-$ *Foxp3*^{Cre} population ($n = 9$ mice per group). **b**, Representative histology of gastrointestinal tract—duodenum, jejunum, ileum and colon (left to right) ($n = 3$ mice per group). There were no significant lesions observed or differences in inflammation, epithelial changes, or mucosal architecture between the two groups (H&E, original magnification, ×100). Scale bar, 50 µm. **c**, Histopathology score in the small intestine and colon of fT$_{reg}$ knockout and control mice. The severity and extent of inflammation and epithelial changes as well as mucosal architecture were each graded on a score of 1 (minimal) to 5 (severe) and added to obtain an overall score over 20. There were minimal inflammatory changes with no significant differences between groups. **d**, Multiplex inflammation panel of serum from fT$_{reg}$ knockout and control mice ($n = 4$ pooled samples (3 mice per sample) per group). Data are mean ± s.e.m. $*P < 0.05$, $***P < 0.001$, Student's *t*-test.



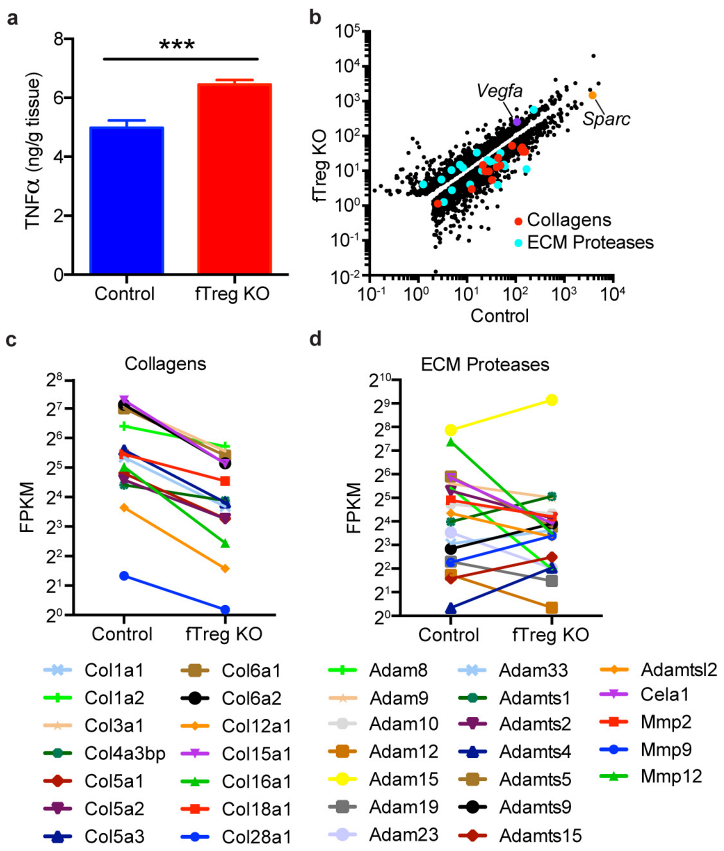
Extended Data Figure 4 | Weight-matched cohorts' body weights and adipocyte size frequency in VAT of aged control and fT_{reg} knockout mice. **a**, Body weights of fT_{reg} knockout and control male mice used in weight-matched metabolic studies in young (12 week; control $n=9$; fT_{reg} KO $n=9$), aged (36 week; control, $n=9$ mice; fT_{reg} KO, $n=11$ mice) and obese (diet-induced obese, 12 weeks of HFD starting at 12 weeks; control $n=10$; fT_{reg} KO $n=10$) settings. **b**, Frequency of small (0–5,000 μm^2), medium (5,000–10,000 μm^2) and large (>10,000 μm^2) adipocytes in VAT of aged control and fT_{reg} knockout mice ($n=3$ mice per group, 850 adipocytes counted from control mice, 269 adipocytes counted from fT_{reg} knockout adipose). Data are mean ± s.e.m.



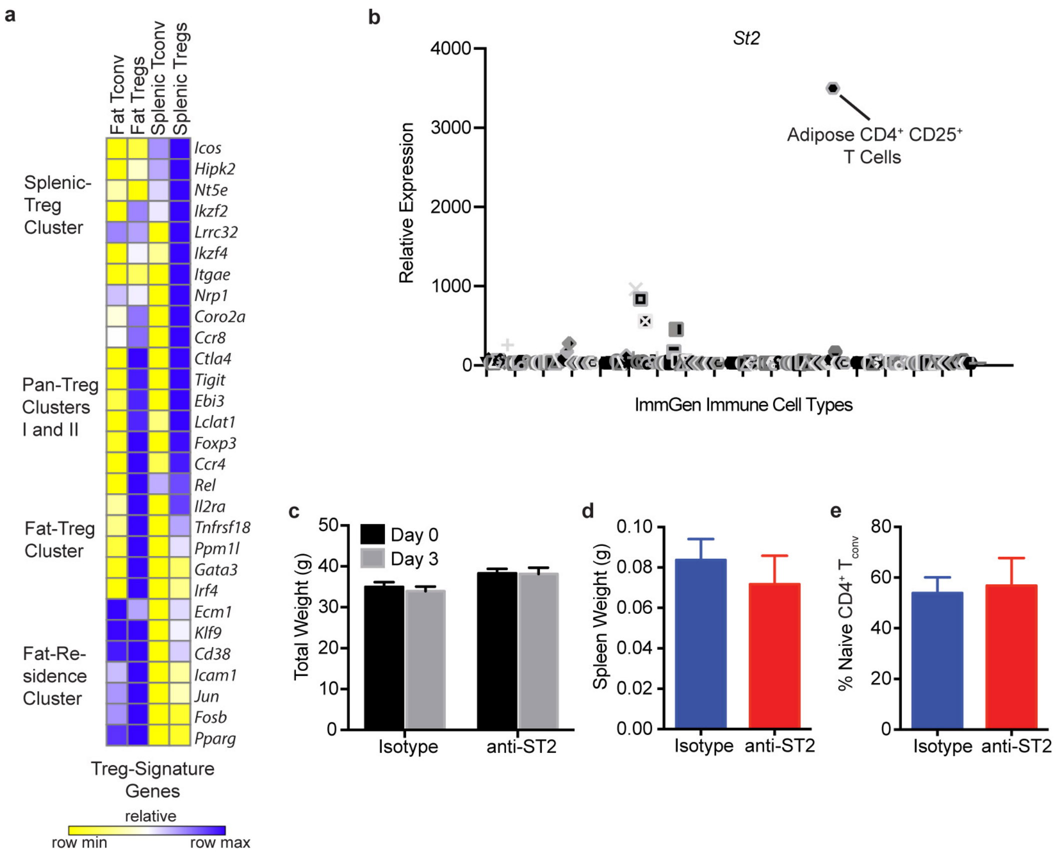
Extended Data Figure 5 | VAT AIPs of obese fT_{reg} knockout and control mice. a, AIPs of diet-induced obese (16 weeks high fat diet started at 12 weeks) control ($n=6$ mice) and fT_{reg} knockout ($n=8$ mice) male mice depicting immune cell abundance, expressed as percentage of CD45.2$^+$ cells. Entirety of immune compartment (top) is further divided into pan-macrophage (middle) and non-macrophage (bottom) pie charts. **b**, Immune cell abundance between fT_{reg} knockout and control mice, expressed as cells per gram of VAT ($n=9$ mice per group). Data are mean ± s.e.m.



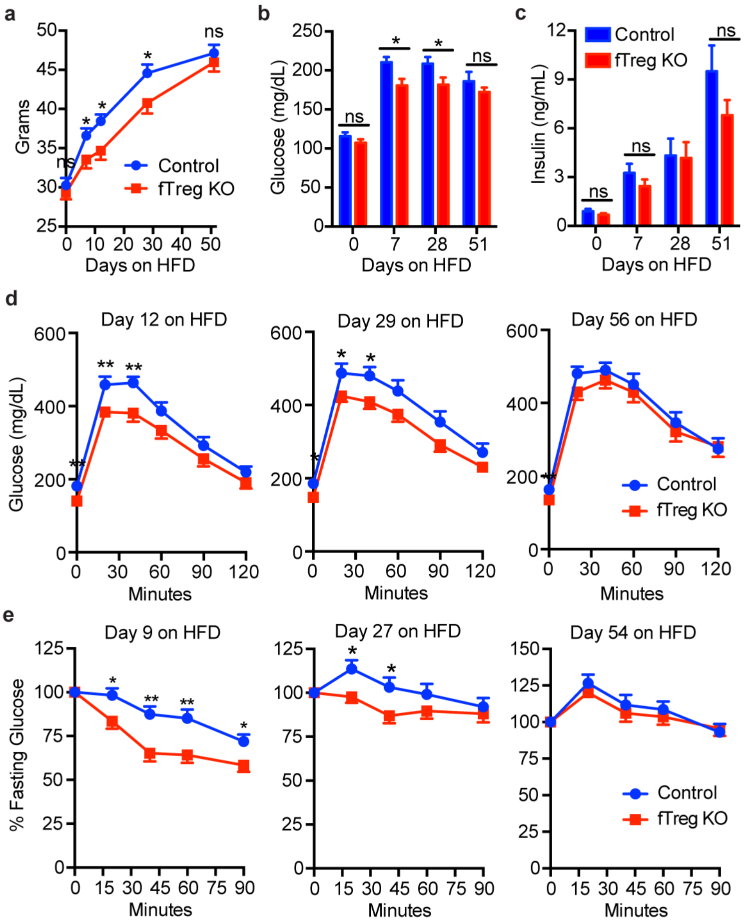
Extended Data Figure 6 | fT_{reg} cells are dispensable for TZDs to exert their therapeutic insulin-sensitizing effect. a, Scheme used for longitudinal interventional study of control and fT_{reg} knockout mice which indicates when particular assays were conducted and whose results are described in **b**–**g**, in which rosiglitazone (Rosi) was introduced in diet after firmly establishing obesity with a HFD alone for 12 weeks ($n = 8$ mice per group). **b**, Cohort weights during course of study. Black arrow indicates introduction of rosiglitazone to the diet. **c**, Homeostatic model assessment of IR (HOMA-IR). **d**, **e**, Glucose tolerance test (**d**) and glucose excursions of glucose tolerance test (**e**) described as area under curve (AUC). **f**, **g**, Insulin tolerance test (**f**) and bar-graph quantitation of relative serum glucose decrease during insulin tolerance test (**g**) described as area above curve (AAC). **h**, Scheme used for parallel prophylactic study of control and fT_{reg} knockout mice, the results of which are described in **i**–**l**, in which mice were placed on a HFD or HFD with rosiglitazone for 12 weeks ($n = 8$ mice per group). **i**, Cohort weights at end of study. **j**, HOMA-IR. **k**, **l**, Glucose and insulin tolerance tests of control (**k**) or fT_{reg} knockout (**l**) mice fed HFD or HFD with rosiglitazone. **m**, Scheme used to determine temporal relationship of TZD-induced fT_{reg} expansion and TZD-induced insulin-sensitization in wild-type mice, the results of which are described in **n**–**q**, where mice were fed HFD or HFD with rosiglitazone for up to 11 weeks ($n = 10$ mice per group, 5 mice of each group were euthanized at 5 weeks after diet introduction and remaining 5 mice were euthanized at 11 weeks). **n**, HOMA-IR at 4 weeks. **o**, **p**, Glucose (**o**) and insulin (**p**) tolerance tests at 5 weeks. **q**, Relative fT_{reg} cell enrichment of mice fed HFD with rosiglitazone versus mice fed HFD alone at 5 and at 11 weeks. Data are mean ± s.e.m. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, Student's t-test.



Extended Data Figure 7 | Increased TNFα levels and gene expression pattern of aged fT_{reg} knockout adipose tissue is consistent with an improved adipose remodelling capacity. a, TNFα levels quantified by ELISA of whole adipose lysate (~40 weeks, $n = 6$ per group). **b**–**d**, FPKM values of all differentially expressed genes (**b**), differentially expressed collagens (**c**) and differentially expressed extracellular matrix proteases (**d**) in VAT from aged fT_{reg} knockout and control mice (~40 weeks, $n = 3$ mice per group). Data are mean ± s.e.m. $^{***}P < 0.001$, Student's *t*-test.



Extended Data Figure 8 | fT_{reg} cell gene expression and depletion with anti-ST2 antibody treatment. **a**, Expression of several canonical T_{reg} cell genes across fat and splenic T_{reg} and fat and splenic T_{conv} cells. Cells were pooled from 3 and 4 mice before isolating RNA for subsequent RNA-Seq analysis. **b**, Expression of *St2* across all haematopoietic cells catalogued in the ImmGen database. Position of adipose $CD4^+$ $CD25^+$ T cells is marked. **c**, Total weight before beginning course of anti-ST2 or isotype control antibodies (day 0) and after terminal analysis (day 3) ($n = 4$ mice per group). **d**, **e**, Spleen weight (**d**) and percentage of splenic naive $CD4^+$ T cells as defined by $CD62^{hi}$ $CD44^{lo}$ relative to total splenic $CD45^+$ $CD4^+$ $CD25^-$ T cell population (**e**) of mice after terminal analysis (day 3, $n = 4$ mice per group). Data are mean ± s.e.m.



Extended Data Figure 9 | Aged fT$_{reg}$ knockout mice are resistant to short-term, but not persistent, HFD-induced weight gain and IR. **a–e**, Aged control and fT$_{reg}$ knockout mice were placed on HFD and monitored throughout course of diet for weight (**a**), fasting glucose levels (**b**), fasting serum insulin levels (**c**), performance on glucose tolerance test (**d**), and on insulin tolerance test (**e**). (Control, $n = 10$; fT$_{reg}$ KO, $n = 11$; mice were aged 27–29 weeks and weight-matched before HFD was introduced.) Data are mean ± s.e.m. ns, non-significant, $*P < 0.05$, $**P < 0.01$, Student's *t*-test.

Extended Data Table 1 | Antibodies used to identify the given immune cell type molecularly

Immune Cell Type	Molecular Identification Scheme
TCRγδ	$CD45.2^{+}$ $F4/80^{-}$ $CD3^{+}$ $TCR\beta^{-}$ $TCR\gamma^{+}$
$CD8^{+}$	$CD45.2^{+}$ $F4/80^{-}$ $CD3^{+}$ $TCR\beta^{+}$ $CD4^{-}$ $CD8^{+}$
Treg $CD4^{+}$	$CD45.2^{+}$ $CD4^{+}$ $CD25^{+}$ $Foxp3^{+}$
Naive $CD4^{+}$	$CD45.2^{+}$ $CD4^{+}$ $CD25^{-}$ $Foxp3^{-}$ $CD62L^{hi}$ $CD44^{lo}$
Activated $CD4^{+}$	$CD45.2^{+}$ $CD4^{+}$ $CD25^{-}$ $Foxp3^{-}$ $CD62L^{lo}$ $CD44^{hi}$
NKT	$CD45.2^{+}$ $NK1.1^{+}$ $TCR\beta^{+}$
NK	$CD45.2^{+}$ $NK1.1^{+}$ $TCR\beta^{-}$
B	$CD45.2^{+}$ $NK1.1^{-}$ $CD19^{+}$
Eosinophil	$CD45.2^{+}$ $F4/80^{+}$ Siglec-F^{+}
Neutrophil	$CD45.2^{+}$ $F4/80^{-}$ $CD11c^{-}$ $CD11b^{+}$ $Ly6G^{+}$
M2 ATM	$CD45.2^{+}$ $F4/80^{+}$ $CD11c^{med}$ $CD206^{+}$
M1 ATM	$CD45.2^{+}$ $F4/80^{+}$ $CD11c^{hi}$ $CD206^{-}$
DN (Double-negative) ATM	$CD45.2^{+}$ $F4/80^{+}$ $CD11c^{-}$ $CD206^{-}$

doi:10.1038/nature15740

Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples

Wenfei Jin[1]*, Qingsong Tang[1]*, Mimi Wan[2], Kairong Cui[1], Yi Zhang[1,3], Gang Ren[1,4], Bing Ni[3], Jeffrey Sklar[2], Teresa M. Przytycka[5], Richard Childs[6], David Levens[7] & Keji Zhao[1]

DNase I hypersensitive sites (DHSs) provide important information on the presence of transcriptional regulatory elements and the state of chromatin in mammalian cells[1–3]. Conventional DNase sequencing (DNase-seq) for genome-wide DHSs profiling is limited by the requirement of millions of cells[4,5]. Here we report an ultrasensitive strategy, called single-cell DNase sequencing (scDNase-seq) for detection of genome-wide DHSs in single cells. We show that DHS patterns at the single-cell level are highly reproducible among individual cells. Among different single cells, highly expressed gene promoters and enhancers associated with multiple active histone modifications display constitutive DHS whereas chromatin regions with fewer histone modifications exhibit high variation of DHS. Furthermore, the single-cell DHSs predict enhancers that regulate cell-specific gene expression programs and the cell-to-cell variations of DHS are predictive of gene expression. Finally, we apply scDNase-seq to pools of tumour cells and pools of normal cells, dissected from formalin-fixed paraffin-embedded tissue slides from patients with thyroid cancer, and detect thousands of tumour-specific DHSs. Many of these DHSs are associated with promoters and enhancers critically involved in cancer development. Analysis of the DHS sequences uncovers one mutation (chr18: 52417839G>C) in the tumour cells of a patient with follicular thyroid carcinoma, which affects the binding of the tumour suppressor protein p53 and correlates with decreased expression of its target gene *TXNL1*. In conclusion, scDNase-seq can reliably detect DHSs in single cells, greatly extending the range of applications of DHS analysis both for basic and for translational research, and may provide critical information for personalized medicine.

We developed a circular carrier DNA-mediated sequencing method, called scDNase-seq, to analyse genome-wide DHSs in a few cells or even single cells (Fig. 1a). Application of scDNase-seq to NIH3T3 cells

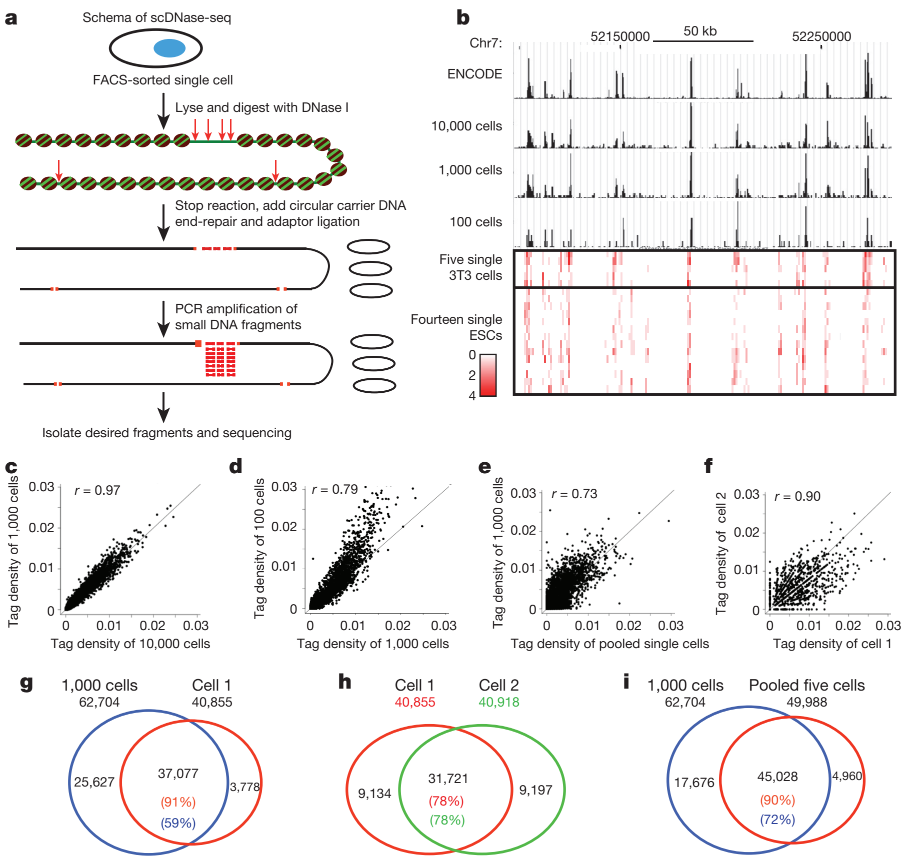


Figure 1 | Genome-wide detection of DHSs in single cells using scDNase-seq. a, Schema of scDNase-seq. Fluorescence-activated cell sorting (FACS)-sorted single cells were digested with DNase I, followed by end-repair, adaptor ligation, polymerase chain reaction (PCR) amplification in the presence of circular carrier DNA and sequencing. **b**, Genome Browser displays showing the DHS in ENCODE data and small cell number scDNase-seq data (black tracks). The red tracks show scDNase-seq read densities in DHSs of 5 single NIH3T3 cells and 14 single mouse ESCs. **c**–**f**, Scatter plots showing the tag density correlation of DHSs between two libraries. Each dot represents one or more DHSs. **g**–**i**, Venn diagrams showing the significant overlaps of DHSs between two libraries.

[1]Systems Biology Center, Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [2]Department of Pathology, Yale University School of Medicine, New Haven, Connecticut 06520, USA. [3]Institute of Immunology, Third Military Medical University of the People's Liberation Army, Chongqing 400038, China. [4]College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China. [5]Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20892, USA. [6]Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [7]Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.
*These authors contributed equally to this work.

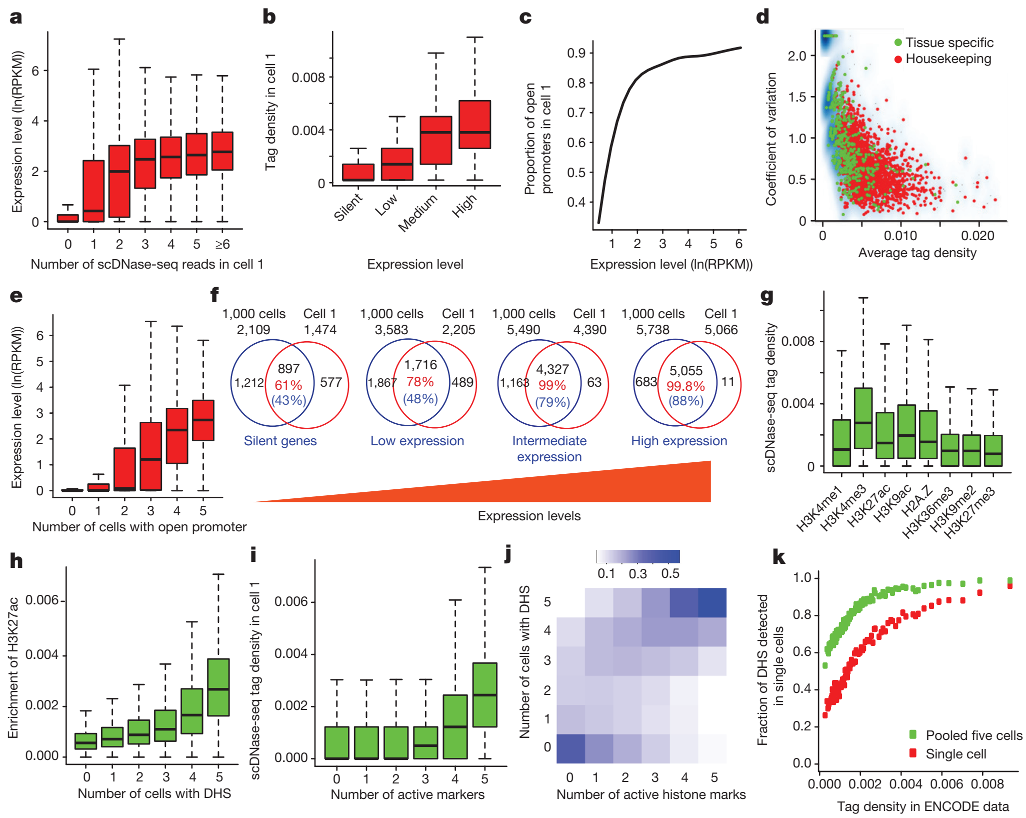


Figure 2 | Detectability of single-cell DHSs is positively correlated with gene expression and number of active histone modifications. **a**, Number of tags within ±1 kilobase (kb) of TSSs correlated with higher gene expression in single cell number 1. **b**, scDNase-seq tag density in single cell 1 is positively correlated with gene expression in a population of cells. **c**, The proportion of open promoters detected by scDNase-seq in single cell 1 is positively correlated with gene expression. **d**, Housekeeping genes (red) show higher tag density and lower variation than tissue-specific genes (green). **e**, Genes with open promoter in more single cells are associated with higher expression levels. **f**, The percentage of overlaps between DHSs detected in 1,000 NIH3T3 cells and single cell 1 positively correlated with gene expression. The total number of genes with DHSs for each group is indicated outside the Venn diagrams. The numbers in red indicate the percentages of DHSs detected in single cell 1 that overlapped with the DHSs detected in 1,000 cells. **g**, Active histone modifications (H3K4me1, H3K4me3, H3K9ac, H3K27ac and H2A.Z) are associated with higher scDNase-seq tag density than the repressive H3K27me3 and H3K9me2 modifications in single cells. **h**, The H3K27ac level effectively predicts the detectability by scDNase-seq. **i**, The scDNase-seq density in cell 1 correlated with the number of histone active modifications. **j**, The detection of DHSs across multiple single cells is positively correlated with the number of histone modifications. **k**, The DHS detectability in single cells is correlated with the tag density of DHS peaks in the ENCODE data.

generated DHS profiles of 10,000, 1,000, 100 and even single cells comparable to that of mouse ENCODE data obtained from 10 million to 20 million cells (Fig. 1b). On average, about 317,000 unique scDNase-seq reads and 38,000 DHSs were detected per single cell. Although the numbers of mapped reads and DHSs decrease as the cell numbers decrease, the enrichments of reads in DHSs in different single cells were very similar (23–26% of reads in DHS regions), despite minor differences (Supplementary Tables 1–3). Scatter plot analysis indicated that the DHSs from 10,000, 1,000 and 100 cells are as reproducible as the ENCODE data (Fig. 1c, d and Extended Data Fig. 1a–c). The pooled DHSs of five single NIH3T3 cells were significantly correlated with those of 1,000 cells (Fig. 1e). We also observed high correlation of DHSs between single cells (Fig. 1f and Extended Data Fig. 1d–l). Venn diagrams showed that at least 90% of DHSs in single cells could be detected in the 1,000-cells data (Fig. 1g and Extended Data Fig. 2a–d). Large fractions (41–82%) of DHSs were shared between two single cells (Fig. 1h and Extended Data Fig. 2e–m). Although only 35–59% of the DHS in the 1,000-cells data were detected in each single cell (Fig. 1g and Extended Data Fig. 2a–d), detectability increased to 72% when the five single cells were pooled (Fig. 1i), suggesting single-cell-specific DHSs contribute to the total number of DHSs detected in a population of cells.

The false discovery rates (FDRs) of single-cell libraries were 11–13% (Supplementary Table 2) when one scDNase-seq tag was detected within a DHS region, suggesting that even detection of one tag is likely to represent a true DHS. Indeed, transcription start sites (TSSs) with one tag exhibited significantly higher expression levels than those without any tag (Fig. 2a and Extended Data Fig. 3a–d). The tag number at TSSs positively correlated with expression levels when the number was low (zero to three tags), but expression levels did not significantly change when the number was high (more than three tags) (Fig. 2a and Extended Data Fig. 3a–d), indicating that the gene expression was no longer limited by accessibility once the promoter had become accessible. As expected, the tag density at TSSs in each single cell correlated with gene expression levels measured in a population of cells (Fig. 2b and Extended Data Fig. 3e–h), and almost all promoters of highly expressed genes were accessible in each single cell (Fig. 2c and Extended Data Fig. 3i–l). Consistent with these observations, the tag densities at housekeeping genes were higher and variations lower than those at tissue-specific genes (Fig. 2d). The number of cells where a promoter exhibited DHS correlated with its gene expression: the genes with DHSs across all five single cells had the highest expression level (Fig. 2e). Further analysis showed that the genes with the lowest cell-to-cell variation at promoters were significantly enriched in basic cell functions such as transcription, cell cycle and RNA processing (Supplementary Table 4). The genes with the highest cell-to-cell variation were significantly enriched in metal ion binding (Supplementary Table 5).

Next we examined the fraction of overlapping open promoters where DHS was detected in either 1,000 cells or one single cell in different expression groups. The analysis revealed that although only 58–61% of

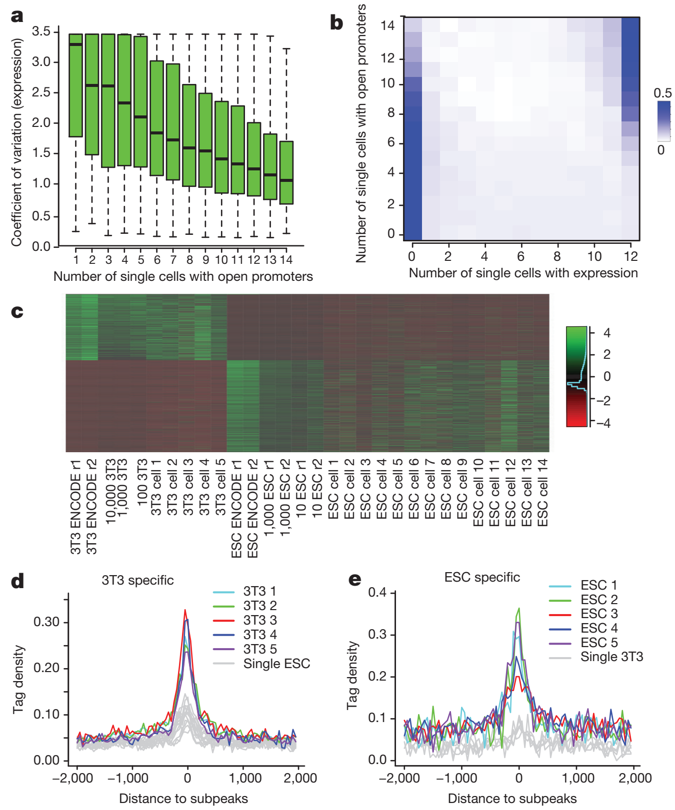


Figure 3 | Single-cell scDNase-seq DHS data can predict cell-specific enhancers. a, Genes with DHSs in fewer single cells (*x* axis) exhibit much higher variation of gene expression across different single cells (*y* axis). **b**, Genes with DHSs in fewer single cells (*y* axis) are expressed in fewer single cells (*x* axis). **c**, The NIH3T3- and ESC-specific DHSs identified in single cells showed expected cell specificity in all libraries. **d**, The subpeaks of NIH3T3-specific super-enhancers show much higher tag density in NIH3T3 cells than that in ESCs. **e**, The subpeaks of ESC-specific super-enhancers show much higher tag density in ESCs than that in NIH3T3 cells.

the open promoters overlapped in the silent gene group, 98–99.9% of the open promoters in intermediately and highly expressed gene groups detected in a single cell overlapped with those detected in 1,000 cells (Fig. 2f and Extended Data Fig. 4), indicating that the DHSs of active genes can be consistently detected in single cells.

Compared with promoter/proximal DHSs, distal DHSs showed lower tag density, higher cell-to-cell variation and noise (Extended Data Fig. 5a–d). Nevertheless, distal DHSs in single cells were clearly enriched in active histone modifications (H3K4me1, H3K4me3, H3K9ac, H3K27ac and H2A.Z) but not repressive ones (H3K36me3, H3K9me2 and H3K27me3) (Fig. 2g), which is consistent with the scenario at the population level[6–11] and validated our single-cell assay. Interestingly, DHS detectability in single cells correlated with the degree of enrichment of the active histone modification (Fig. 2h and Extended Data Fig. 5e–h), and correlated with the number of active marks at the DHSs (Fig. 2i and Extended Data Fig. 5i–l). The vast majority of DHSs were detected across all five single cells when five active histone modification marks were present, whereas DHSs exhibited in variable number of cells when only one or two active marks were present (Fig. 2j). These results indicate that DHS at enhancers are variable between different cells and provide strong evidence that multiple active histone modifications strongly correlated with chromatin accessibility across different single cells.

We compared the DHSs detectability in single cells with the tag density of DHSs from 1,000 cells or 20 million cells. The results indicated the detectability of DHSs in single cells positively correlated with the tag density from the library by a large number of cells (Fig. 2k and Extended Data Fig. 5m). We hypothesized that strong DHSs would be present in all the cells and weak DHSs in only a fraction. If this were the case, more strong DHSs and fewer weak DHSs should be detected within one single cell. Indeed, 80–90% of the strong DHSs were detected whereas only 20–30% of weak DHSs were detected in single cells (Fig. 2k and Extended Data Fig. 5m). Another prediction from this hypothesis was that relatively fewer strong DHSs and more weak DHSs would be additionally detected as we added up single cells. Pooling the five single cells indeed showed the fraction of detected weak DHSs was doubled, whereas the fraction of detected strong DHSs only increased by a small percentage (Fig. 2k and Extended Data Fig. 5m).

The variation of DHSs among single cells within a 'homogenous' population is reminiscent of the well-known phenomenon of variation of gene expression among single cells[12]. To study their relationship, we constructed 14 single embryonic stem-cell (ESC) scDNase-seq libraries (Supplementary Tables 6–8). Comparison with single-cell RNA sequencing (RNA-seq) data[13] revealed that tag density and variation at TSS of single-cell scDNase-seq indeed correlated with that of single-cell gene expression (Extended Data Fig. 6a, b). Furthermore, the genes with DHSs in fewer single cells showed high variation of expression and were expressed in fewer single cells (Fig. 3a, b). These results further indicate that the cell-to-cell variations of single-cell DHSs are predictive of gene expression. Consistent with this notion, we found a significantly higher correlation between the technical repeats compared with that of two non-technical repeat libraries (Extended Data Fig. 6c, d). The Gene Ontology (GO) terms enriched among genes with the lowest and the highest cell-to-cell variation in the 14 ESCs were consistent with that in single 3T3 cells (Supplementary Tables 9 and 10).

We next identified 1,735 NIH3T3-specific DHSs and 2,180 ESC-specific DHSs using the 5 NIH3T3 and 14 ESC single-cell scDNase-seq libraries. Heat map showed these cell-specific DHSs displayed expected cell specificity in all of the libraries (Fig. 3c). The cell-specific DHSs were highly correlated with cell-specific gene expression (Extended Data Fig. 6e, f) and enriched in distinct biological functions (Extended Data Fig. 6g, h). Super-enhancers play a key role in regulating expression of critical cell-specific genes[14–16]. We identified 275 NIH3T3-specific and 231 ESC-specific super-enhancers and compared their single-cell scDNase-seq tag densities. The subpeaks of 3T3-specific super-enhancers were associated with a substantially higher tag density in single 3T3 cells than that in ESCs, and vice versa (Fig. 3d, e), indicating that single-cell DHSs can help predict super-enhancers.

Chromatin defects underlie various diseases including cancers[17]. Profiling genome-wide chromatin accessibility in cells from patients, which are often limiting in numbers, would be clinically invaluable. We applied scDNase-seq to cells dissected from a follicular thyroid carcinoma (FTC) sample on formalin-fixed paraffin-embedded (FFPE) slides (Fig. 4a). DNase I digestion resulted in typical periodic cleavage patterns of nucleosome arrays and read enrichment around TSSs (Fig. 4b and Extended Data Fig. 7a–c). Likewise, the genome browser displays showed peaks (Fig. 4c), suggesting the cells recovered from the FFPE slides retained key chromatin features.

HMGA2 is upregulated in FTC[18,19] and its promoter indeed exhibited higher accessibility in the tumour than that in adjacent normal cells (Fig. 4d). Overall, 1,342 tumour-specific and 2,812 normal-specific DHSs were identified (Extended Data Fig. 8a, b). The genes associated with the tumour-specific DHSs were significantly enriched in the GO biological process terms such as regulation of GTPase activity and response to hypoxia, and pathways such as E-cadherin signalling, RhoA signalling, p53 pathway, RAC1 signalling and MYC transformation (Extended Data Fig. 8). Among these were several known genes such as *TIAM1* and *PIP4K2A* (Extended Data Fig. 9a, b), involved in tumours[20,21]. Interestingly, genes that are characteristic of PAX8–PPARG fusion[22] in FTC were enriched in tumour-specific DHSs (Extended Data Fig. 8f and Supplementary Table 11), even though

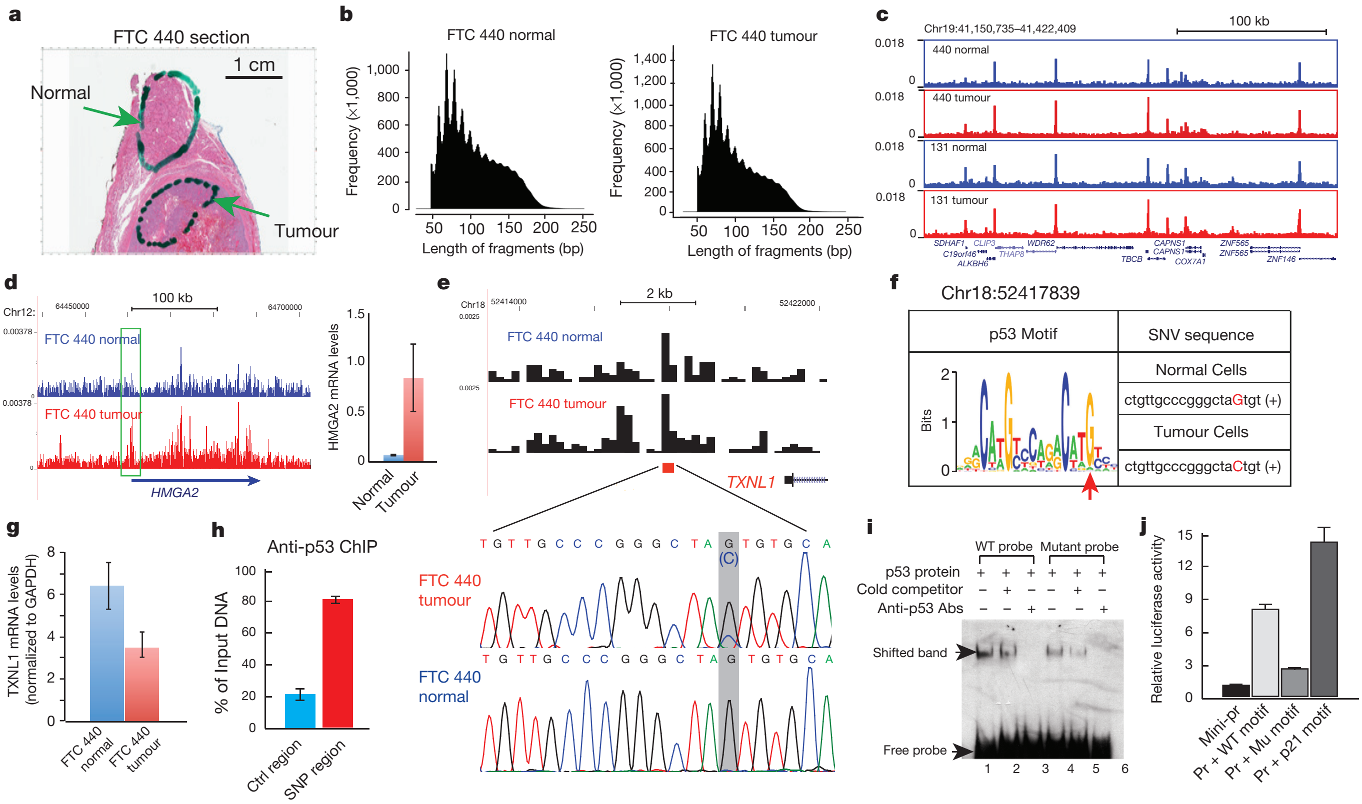


Figure 4 | Application of scDNase-seq to tissue sections from patients with FFPE reveals novel pathophysiological information on thyroid cancers. **a**, Full view of a slide of FTC 440 stained with haematoxylin and eosin. Cells recovered from the highlighted areas were subject to scDNase-seq analysis. **b**, Typical periodic DNase cleavage patterns of nucleosomes were detected both for normal and tumour cells by scDNase-seq. **c**, Genome Browser image displaying the scDNase-seq profiles of the normal (blue) and tumour (red) cells from two thyroid carcinomas 440 and 131. **d**, Genome Browser image showing the increased chromatin accessibility of the *HMGA2* promoter in the tumour cells of FTC 440 (left). Quantitative reverse transcription PCR (qRT–PCR) shows the increased HMGA2 messenger RNA (mRNA) level in the tumour cells (right). **e**, An SNV was identified at a DHS near the 3′ end of the *TXNL1* gene in the tumour cells of FTC 440. The SNV location is indicated by the red square. The SNV was confirmed by Sanger sequencing (highlighted region). **f**, The G to C change in tumour cells negatively impacts the p53 target motif. The SNV in the p53 motif logo is indicated by a red arrowhead. **g**, The G to C change in the tumour cells is correlated with decreased expression of TXNL1 by qRT–PCR analysis. **h**, Protein p53 is bound to the SNV region in a human thyroid cell line by chromatin immunoprecipitation (ChIP)–qPCR analysis. **i**, The G to C change decreases p53 binding affinity *in vitro* by gel shift assay. **j**, The G to C change reduces the activity of the p53 motif to activate a reporter promoter *in vivo*. The p53 motif from the p21 promoter was used as a positive control.

PPARG gene rearrangement was not detected by fluorescence *in situ* hybridization of FTC 440 (data not shown). This suggests that pathways associated with the transcriptional regulation by PAX8–PPARG, but not necessarily the PAX8–PPARG rearrangement itself, are important in mediating follicular thyroid tumorigenesis.

We similarly analysed samples from two more FTC (797 and 957) and one papillary thyroid carcinoma (PTC from patient number 131) samples (Supplementary Table 12). Comparison of the tumour-specific DHSs identified in the three FTC samples revealed very few shared DHSs among all three FTC samples (Extended Data Fig. 10a). The *HMGA2* promoter exhibited a strong DHS in the tumour cells but not in their neighbouring normal cells in FTC from patient number 440, whereas in the other two FTC cases (957 and 797) the promoter showed strong DHSs both in tumour and in normal cells (Extended Data Fig. 10b). Instead, an intronic enhancer showed differential DHSs between the tumour and normal cells (Extended Data Fig. 10b). These results suggest that the mis-regulation of *HMGA2* in the tumour cells may be attributed to different regulatory elements in different patients. Analysis of PTC 131 also identified numerous tumour-cell- and normal-cell-specific DHSs, which are enriched in disease ontologies (Extended Data Fig. 10c). Overall, our results indicate that the vast majority of DHSs are patient-specific, implying that these tumours may arise or progress via different mechanisms in different patients.

To gain further mechanistic insight, we searched for genetic lesions within DHSs in FTC 440 by comparing the DHS sequence between tumour and normal cells. A total of 31 potential single nucleotide variations (SNVs) were identified in the DHS regions, which included both loss of heterozygosity of known SNPs and *de novo* mutations (Supplementary Table 13). We confirmed the *de novo* mutation (chr18:52417839G>C) at a DHS downstream of the thioredoxin-like 1 gene (*TXNL1*) (Fig. 4e). *TXNL1* encodes a regulatory subunit of the human 26S proteasome[23]. Downregulation of TXNL1 is associated with poor prognostic outcomes, aneuploidy in colorectal carcinoma[24] and is implicated in cisplatin-induced apoptosis[25]. Interestingly, the G to C change appears to negatively impact the binding motif of p53 (Fig. 4f) and correlates with significantly decreased expression of *TXNL1* in the tumour cells (Fig. 4g); p53 binds to this DHS in a human thyroid cell line (Fig. 4h). The G to C mutation at this site compromises p53 binding (Fig. 4i) and impairs its ability to activate a reporter promoter (Fig. 4j), suggesting that the G to C change may underlie the decreased *TXNL1* expression in the tumour cells (Fig. 4g). This SNP was not detected in the other three patients (797, 957 and 131). Therefore, our strategy for searching SNVs in relevant DHS regions seems to be a cost-effective alternative to whole-genome sequencing for detecting functionally important mutations in regulatory regions.

Tn5-transposase-mediated detection of chromatin accessibility (scATAC-seq)[26,27] in a large number of single cells has been reported recently. However, the reads per cell generated by scATAC-seq may be too sparse to examine the cell-to-cell variation at individual regulatory regions[26,27]. In comparison, our scDNase-seq detects a much larger

number of DHSs per cell, which provides information on cell-to-cell variations of individual DHSs. scDNase-seq is expected to find its use in multiple settings, such as the analysis of rare cell populations during lineage development and the study of clinical samples with extremely small numbers of cells such as circulating tumour cells, laser-captured cells, core biopsy or fine-needle aspiration samples. Being able to evaluate the chromatin states associated with specific diseases or developmental programs might provide valuable information for developing new diagnostic and therapeutic strategies for these malignancies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 May; accepted 10 September 2015.
Published online 25 November 2015.

1. Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132,** 311–322 (2008).
2. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489,** 75–82 (2012).
3. Sheffield, N. C. *et al.* Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res.* **23,** 777–788 (2013).
4. Song, L. & Crawford, G. E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protoc*. http://dx.doi.org/10.1101/pdb.prot5384 (2010).
5. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature Genet.* **43,** 264–268 (2011).
6. Roh, T. Y., Cuddapah, S. & Zhao, K. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* **19,** 542–552 (2005).
7. Litt, M. D., Simpson, M., Gaszner, M., Allis, C. D. & Felsenfeld, G. Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science* **293,** 2453–2455 (2001).
8. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39,** 311–318 (2007).
9. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107,** 21931–21936 (2010).
10. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470,** 279–283 (2011).
11. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genet.* **40,** 897–903 (2008).
12. Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510,** 363–369 (2014).
13. Sasagawa, Y. *et al.* Quartz-seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* **14,** R31 (2013).
14. Lin, C., Garruss, A. S., Luo, Z., Guo, F. & Shilatifard, A. The RNA Pol II elongation factor Ell3 marks enhancers in ES cells and primes future gene activation. *Cell* **152,** 144–156 (2013).
15. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–319 (2013).
16. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155,** 934–947 (2013).
17. You, J. S. & Jones, P. A. Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell* **22,** 9–20 (2012).
18. Chiappetta, G. *et al.* HMGA2 mRNA expression correlates with the malignant phenotype in human thyroid neoplasias. *Eur. J. Cancer* **44,** 1015–1021 (2008).
19. Belge, G. *et al.* Upregulation of HMGA2 in thyroid carcinomas: a novel molecular marker to distinguish between benign and malignant follicular neoplasias. *Genes Chromosom. Cancer* **47,** 56–63 (2008).
20. Habets, G. G. *et al.* Identification of an invasion-inducing gene, *Tiam-1*, that encodes a protein with homology to GDP-GTP exchangers for Rho-like proteins. *Cell* **77,** 537–549 (1994).
21. Emerling, B. M. *et al.* Depletion of a putatively druggable class of phosphatidylinositol kinases inhibits growth of p53-null tumors. *Cell* **155,** 844–857 (2013).
22. Lui, W. O. *et al.* Expression profiling reveals a distinct transcription signature in follicular thyroid carcinomas with a PAX8-PPARγ fusion oncogene. *Oncogene* **24,** 1467–1476 (2005).
23. Andersen, K. M. *et al.* Thioredoxin Txnl1/TRP32 is a redox-active cofactor of the 26 S proteasome. *J. Biol. Chem.* **284,** 15246–15254 (2009).
24. Gemoll, T. *et al.* HDAC2 and TXNL1 distinguish aneuploid from diploid colorectal cancers. *Cell. Mol. Life Sci.* **68,** 3261–3274 (2011).
25. Ni, P. *et al.* TXNL1 induces apoptosis in cisplatin resistant human gastric cancer cell lines. *Curr. Cancer Drug Targets* **14,** 850–859 (2015).
26. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523,** 486–490 (2015).
27. Cusanovich, D. A. *et al.* Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348,** 910–914 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Cooper and B. Z. Stanton for reading the manuscript, the National Heart, Lung, and Blood Institute DNA Sequencing Core Facility for sequencing the libraries and the National Heart, Lung, and Blood Institute Flow Cytometry Core facility for sorting the cells. The work was supported by Division of Intramural Research, National Heart, Lung, and Blood Institute (to K.Z.), Intramural Research Program of National Library of Medicine (to T.M.P.) and the Center for Cancer Research of the National Cancer Institute (to D.L.) of the National Institutes of Health.

Author Contributions K.Z. conceived the project. Q.T., K.C., M.W. and K.Z. performed the experiments. W.J. analysed the data. Y.Z., G.R., B.N., J.S., T.M.P., R.C. and D.L. contributed to the experiments or data analysis. D.L. and K.Z. directed the project. W.J. and K.Z. wrote the manuscript.

Author Information The scDNase-seq and RNA-seq data sets have been deposited in the Gene Expression Omnibus database with accession number GSE61844. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.Z. (zhaok@nhlbi.nih.gov).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Cell culture and sorting. NIH/3T3 tet-on 3G cells (Clontech, 631197) were cultured in DMEM (Invitrogen, 10566-016) supplemented with 10% FBS (Sigma, F4135-500ML) and 100 U ml^{-1} penicillin–streptomycin (Invitrogen, 15140-122). Mouse ESCs were cultured as described[28]. Single-cell suspension after trypsinization was used for 4′,6-diamidino-2-phenylindole (DAPI) staining immediately before sorting by flow cytometry. Single live cells were sorted and deposited directly into each tube of a PCR strip-tube, which contained 30 μl cell lysis buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Triton X-100).

DNase I digestion and scDNase-seq library preparation. To prevent loss of the extremely small amount of DNase I hypersensitive DNA (<0.1 pg) released by DNase I digestion of single cells, we added a large amount of circular plasmid DNA (30 ng; about 3×10^5 times the amount of the DHS DNA in a single cell) as carrier DNA in the subsequent steps of library preparation. The circular DNA was not compatible with the adaptor ligation and thus could minimize the non-specific amplification by the subsequent PCR. The PCR conditions were optimized to amplify the small fragments (<200 base pairs (bp)) derived from DNase I hypersensitive sites without previous fractionation of these fragments.

For DNase I digestion, 0.2 to 1 unit of DNase I (Roche, 04716728001) was added to the cells and incubated at 37 °C for 5 min. The reaction was stopped by adding 80 μl of stop buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 0.15% SDS, 10 mM EDTA) containing 1 μl of 20 mg ml^{-1} proteinase K and 5 μl of 6 ng μl^{-1} circular carrier DNA. The mixture was incubated at 55 °C for 1 h and DNA purified by phenol–chloroform extraction, followed by precipitation with ethanol in the presence of 20 μg glycogen. The library was prepared using Illumina kits as described[29]. The libraries were amplified using a two-step method to preferentially amplify the small DNA fragments derived from the DNA hypersensitive sites and to reduce non-specific amplification of the carrier DNA. The first amplification was done with index primers with the PCR condition 98 °C for 10 s, 67 °C for 30 s, 72 °C for 30 s for six cycles. After isolation of the desired fragments (160–300 bp) using 2% E-gel (Invitrogen), the second amplification was done with the P5 and P7 primers with the condition 98 °C for 10 s, 68 °C for 30 s, 72 °C for 30 s for 22 cycles. The fragments between 160 and 300 bp were isolated on E-gel and sequenced on Illumina HiSeq 2500.

Recovery of cells from FFPE tissue slides. The anonymized tumour samples from Ambry Genetics, approved by institutional review board and with informed consent, were used in this study. Three cases of thyroid cancer were diagnosed as FTC and one case was diagnosed as papillary thyroid carcinoma. Cells were manually scraped off from the highlighted area of a paraffin slide using a razor blade and resuspended in 150 μl of de-paraffinization solution (Qiagen, 1064343) and incubated at 56 °C for 3 min. After cooling to room temperature (about 25 °C), 150 μl of lysis buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Triton X-100) was added and incubated at 37 °C for 2 h. The cells in the lower layer were transferred to a new tube and digested by DNase I as described above. The formaldehyde cross-linking was reversed by incubating DNA at 65 °C overnight, which was followed by DNA purification and library preparation.

Extraction of total RNAs from cells recovered from FFPE slides, RT–PCR, RNA-seq. Cells recovered from FFPE slides were resuspended in 150 μl of de-paraffinization solution (Qiagen, 1064343) and incubated at 56 °C for 3 min. Total RNA was extracted using an RNA extraction kit from (Qiagen, 73504), following the manufacturer's instructions. After reverse transcription using an oligonucleotide dT primer, the mRNA expression levels of selected genes were analysed using the following gene-specific primers and probes from Applied Biosystems: HMGA2-Hs00171569_ml, TIAM1-Hs01021959_ml, TXNL1-Hs00355488_ml, PIP4K2A-Hs00178197_ml and GADPH-Hs99999905_ml.

The RNA-seq libraries were generated according to established protocols and sequenced on HiSeq 2500 platforms.

Validation of SNVs by Sanger sequencing. The tumour and adjacent normal cells from FFPE slides were recovered and resuspended in 100 μl of 1× TE + 0.1% SDS + 0.2 mg ml^{-1} proteinase K. Following incubation at 65 °C overnight, the genomic DNA was purified using phenol–chloroform extraction and ethanol precipitation. The genomic region containing the potential sequence variation was amplified by PCR using specific primers. The PCR products were then sequenced by Sanger sequencing. Forward primer, AAGCTAAATGAGCAAAATATTCCT; reverse primer, GGGAGGCTGAGGCAGTAGAATCG.

ChIP, electrophoretic mobility shift assay and promoter reporter assays. Chromatin extracts were prepared from a human thyroid cell line (Nthy-ori 3-1 human Cell Line, from Sigma-Aldrich, 90011609). ChIP experiments were performed with p53 antibodies (Santa Cruz Biotechnology, sc-6243X) using established protocols[1]. The ChIP DNA was analysed using qPCR with the following primers: p53 positive forward primer, GTCATGCGATCTTGGCTCACT; reverse primer, CTTGGGAGGCTGAGGCAGTA; probe, CAACCTCCGCCTCCCGGGTTC. Control forward primer, CCCCATGCTGTTCTCGTGATA; reverse primer, GCAAAGGTGAATCAAGGCATCT; probe, TTTATAAGGTTCTCTTCCCCTTTCGCTGGG.

Electrophoretic mobility shift assay (EMSA) experiments were performed using nuclear extracts of HeLa cells transfected with a p53 expression vector (provided by J. Huang). Briefly, the double-stranded oligonucleotide probes (wild-type p53 site, CACTCTGTTGCCCGGGCTAGTGTGCAGT; tumour p53 site, CACTCTGTTGCCCGGGCTACTGTGCAGT; p21 promoter p53 site, CAGGAACAAGTCAAGACATGTTCAGC) were synthesized and labelled with biotin using Biotin 3zeEnd DNA Labelling Kit (Thermo Scientific, 89818). The EMSA assays were conducted by using LightShift Chemiluminescent EMSA Kit (Thermo Scientific, 20148) according to the manufacturer's instructions.

To test the activity of the p53 binding sites to activate a reporter promoter, we cloned the wild type p53 binding motif, the motif with the G to C mutation and the p53 motif from the p21 promoter into the XhoI and BglII upstream of the basal cytomegalovirus promoter driving a luciferase reporter gene (provided by J. Huang). The constructs were transfected into Nthy-ori 3-1 human cell line cells for 2 days and the luciferase activity of whole-cell extracts was measured using a Dual-Luciferase Reporter Assay kit (Promega, E1960). The oligonucleotide sequences used in the reporter constructs were as follows: wild type p53 site, TCGAGCTGTTGCCCGGGCTAGTGTGA; tumour p53 site, TCGAGCTGTTGCCCGGGCTACTGTGA; p21 promoter p53 site, TCGAGGAACAAGTCAAGACATGTTCA.

Data analysis. Data, reads mapping and filtering: in this study, we constructed a total of 38 scDNase-seq libraries including 8 NIH3T3 libraries (Supplementary Table 1), 18 ESC libraries (Supplementary Table 6) and 12 FFPE patient libraries (Supplementary Table 12). Among these libraries, there are 5 NIH3T3 single-cell scDNase-seq libraries and 14 ESC single-cell scDNase-seq libraries. We also prepared eight RNA-seq libraries using cells recovered from the FFPE tissue section slides of FTC 440 (Supplementary Table 12). In addition to the scDNase-seq and RNA-seq libraries prepared in this study, we integrated the histone modification ChIP-seq data of NIH3T3 from our previous study[30]. We also downloaded the DNase-seq data of NIH3T3 cells and ESCs from mouse ENCODE project[31]. Reads of DNase-seq/scDNase-seq/ChIP-seq were mapped to the mouse genome (mm9) or human genome (hg18) using Bowtie2 (ref. 32). Iterative alignment, in which the unmapped reads were trimmed 5 bp and were re-aligned until reads were less than 26 bp, were conducted for small cell number scDNase-seq libraries and single-cell scDNase-seq libraries. The reads with mapping quality (MAPQ) ≤ 10 or redundant reads that mapped to the same location with the same orientation were removed from further analysis in each library. The mappability of 1,000-cells scDNase-seq libraries to the mouse or human genome was about 40% whereas that of the single-cell scDNase-seq libraries was about 2% owing to non-specific amplification of carrier DNA. The tag density at each bin of 200 bp was calculated by normalizing the number of reads in the bin to the total number of reads in the library and the bedgraphs were uploaded to the UCSC Genome Browser.

Peak calling for DNase-seq/scDNase-seq and correlation between different libraries: the DHSs in mouse ENCODE DNase-seq data and small cell number scDNase-seq data were identified using model-based analysis of ChIP-seq (MACS)[33] by setting a P value to 1×10^{-5}. The peaks identified in the ENCODE data were extended ±1 kb from the summit of the peak if the peak size was <2 kb and overlapping peaks were merged. Then the number of reads in each DHS for all DNase-seq and scDNase-seq libraries was counted. The tag density at each DHS was calculated by normalizing the number of reads in the DHS to the total number of reads in the library (possibility of a tag located on a base-pair per million reads). The Pearson product-moment correlation coefficient (r) of tag densities at genome-wide DHS between two libraries was calculated to indicate the correlation between different scDNase-seq libraries. For single-cell libraries, the reads out of the defined DHS regions were filtered and the number of reads in each 1,000-bp bin was counted to generate the single-cell heat map (Fig. 1b). Any DHS region in a single cell with a reads located in was treated as open access thus a DHS in this single cell. For the pooled five single cells, any DHS region with at least two reads located in was treated as the DHS in the pooled five single cells.

The FDR of the DHS detected in single cells: in an NIH3T3 single-cell scDNase-seq library, the total number of observed DHSs and false positive (type I error) DHSs were denoted by N_{DHS} and N_{FP}, respectively. On the other hand, any reads that located out of the DHSs detected in ENCODE data must have been caused by noise generated during library preparation. The noise level (σ) should be the total number of reads that located out of the DHS in ENCODE data dividing by total length of the regions that are not DHS. The number of false positive DHSs

should be the genome-wide noise level (σ) multiplying by the total length of the DHS. Thus, the FDR should be the number of false positive DHSs dividing by all the detected DHSs in single cell:

$$\mathrm{FDR} = \frac{N_{\mathrm{FP}}}{N_{\mathrm{DHS}}} = \frac{\sigma \times \mathrm{length}_{\mathrm{DHS}}}{N_{\mathrm{DHS}}}$$

On the basis of this formula, we calculated the FDR for each NIH3T3 and ESC single-cell scDNase-seq library (Supplementary Tables 2 and 7).

Differentially expressed genes and tissue-specific genes: the reads from RNA-seq libraries were mapped to the mouse genome (mm9) or human genome (hg18) using Bowtie2 (ref. 32). The gene expression level was measured by reads per kilobase per million mapped reads (RPKM) and number of reads in each gene. The cell-specific genes between ESC and NIH-3T3 were identified using EdgeR (FDR $<$ 0.05; fold change $>$ 1.5 or greater than two-thirds)[34].

We used the tissue specificity index τ (ref. 35) to measure the tissue specificity of each gene, which is defined as the heterogeneity of its expression level across all the tissues. Assuming there are n tissues, the expression level of a gene in the jth tissue is $E(j)$ and the highest expression level of the gene across all tissues is $E_{\max}$. Thus τ is calculated by

$$\tau = \sum_{j=1}^{n} \left(1 - \frac{\log_2 E(j)}{\log_2 E_{\max}}\right) / (n-1)$$

The values of τ range from 0 to 1, with higher values indicating higher variation of expression across tissues and thus higher tissue specificity, whereas lower values indicate lower variation of expression across tissues. The genes with the lowest τ could be considered as housekeeping genes. In this study, we calculated τ on the basis of gene atlas data from bioGPS. The 2,000 genes with the highest τ and the 2,000 genes with the lowest τ were treated as the tissue-specific genes and housekeeping genes, respectively.

The histone modification ChIP-seq data and peak calling: since the peaks of some histone marks such as H3K36me3 and H3K27me3 are very broad, we identified the tag-enriched peaks using SICER[36], which takes advantage of the enrichment information from neighbouring bins to identify spatial clusters of signals that are unlikely to appear by chance. We set the window size to 200 bp and FDR = 0.01 for each histone modification ChIP-seq library, while we set the gap to 200 bp for H3K4me3, H3K9ac; 400 bp for H2A.Z; and 600 bp for the H3K4me1, H3K9me2, H3K27ac and H3K27me3. We calculated the tag densities of each active histone modification peak and identified whether the peak was a DHS in each single cell to find whether the enrichment of an active histone mark was correlated with the number of cells with DHS at the same locus. We calculated the tag densities of each single-cell scDNase-seq library at each DHS and examined whether a DHS co-occurred with these active histone modifications to find whether the chromatin accessibility in each single cell was correlated with the number of histone modifications in the same locus. Two peaks from different libraries were considered a co-occurrence if the overlapped region accounted for $>$10% of the length of a peak.

Reads around promoters and subpeaks of super-enhancers: the RefSeq genes (mm9 and hg18) were downloaded from the UCSC Genome Browser database. The regions $\pm$1 kb around the TSS were treated as promoters in this study. The number of scDNase-seq reads located in a promoter was used to measure the chromatin accessibility of the promoter. We searched the super-enhancer in NIH3T3 via ROSE[15] on the basis of H3K27ac ChIP-seq and scDNase-seq data, respectively. We obtained a total of 275 high-confidence super-enhancers in NIH3T3 by identifying super-enhancers shown both in H3K27ac and in DNase-seq data. In addition, the 231 super-enhancers in ESCs reported in ref. 15 were used in this study. Subpeaks in super-enhancers were identified by MACS[33] and average read densities around these subpeaks of super-enhancers were calculated.

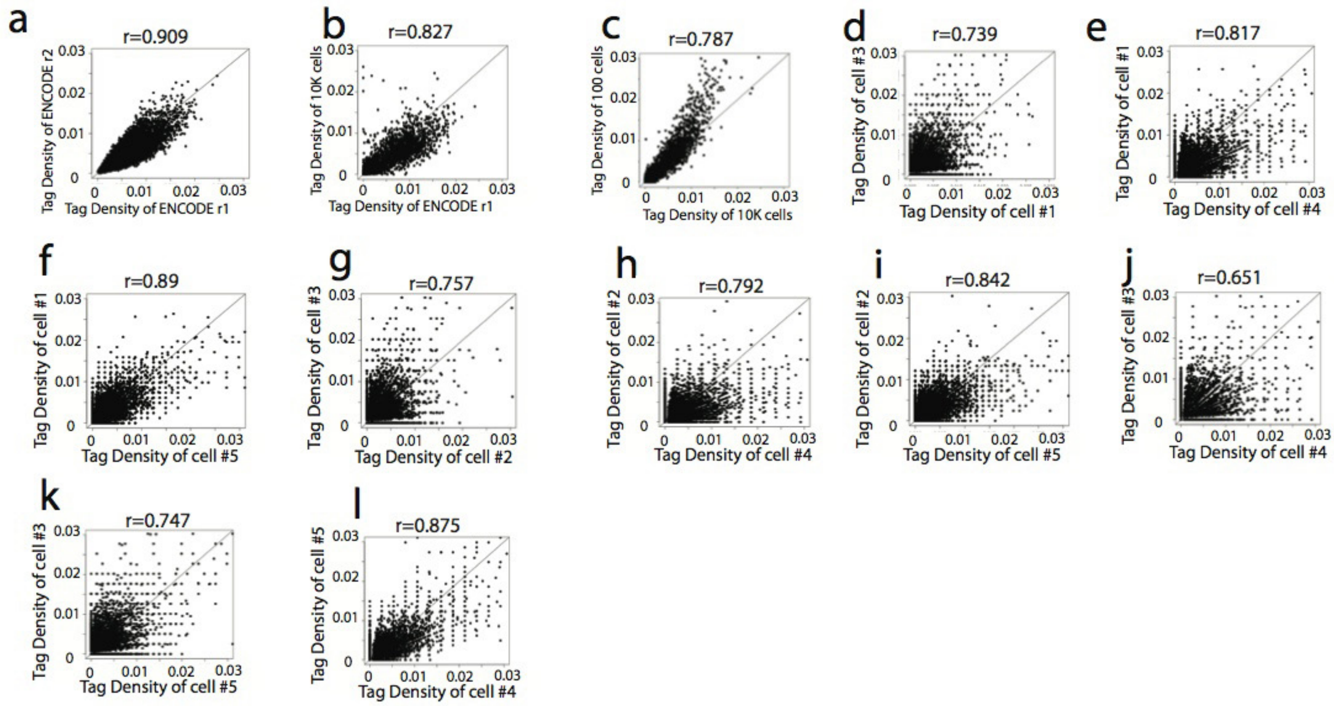
Single-cell-specific DHSs and gene set enrichment analysis: the number of reads located each DHS detected in ENCODE data in each NIH3T3 cell and ESC was counted. To examine whether the chromatin accessibility between NIH3T3 cells and ESCs was significantly different, a Wilcoxon signed-rank test was performed on the number of reads in the 5 NIH3T3 cells and 14 ESCs at each DHS. A DHS was active (indicated by 1) in a single cell if there was one or more than one reads located in the DHS region in the cell, while the alternative was not active (indicated by 0). Fisher's exact test on each locus was performed on the number of cells with active DHSs and the number of cells without active DHSs between the 5 NIH3T3 cells and 14 ESCs. The DHSs with $P < 0.05$ both by Wilcoxon test and by Fisher's test were treated as cell-type specific. Finally, we identified 1,735 single-cell NIH3T3-specific DHSs and 2,180 single-cell ESC-specific DHSs. We used gene set enrichment analysis[37] to determine whether the genes in the vicinity of the single-cell-specific DHSs showed statistically significant differences between NIH3T3 cells and ESCs on the basis of the gene expression data.

Gene ontology of single-cell NIH3T3-specific and ESC-specific DHSs: to predict the function of single-cell NIH3T3-specific or ESC-specific DHSs, we performed gene ontology analysis using GREAT[38] with the 1,735 NIH3T3-specific and 2,180 ESC-specific DHSs. It is clear that the single-cell ESC-specific DHSs are enriched with stem cell development and differentiation genes, and the single-cell NIH3T3-specific DHSs are enriched with genes with different functions (Extended Data Fig. 6g, h). These results indicate that the ESC-specific and NIH3T3-specific DHSs identified in the single-cell scDNase-seq libraries predict important enhancers critical for tissue-specific gene expression.

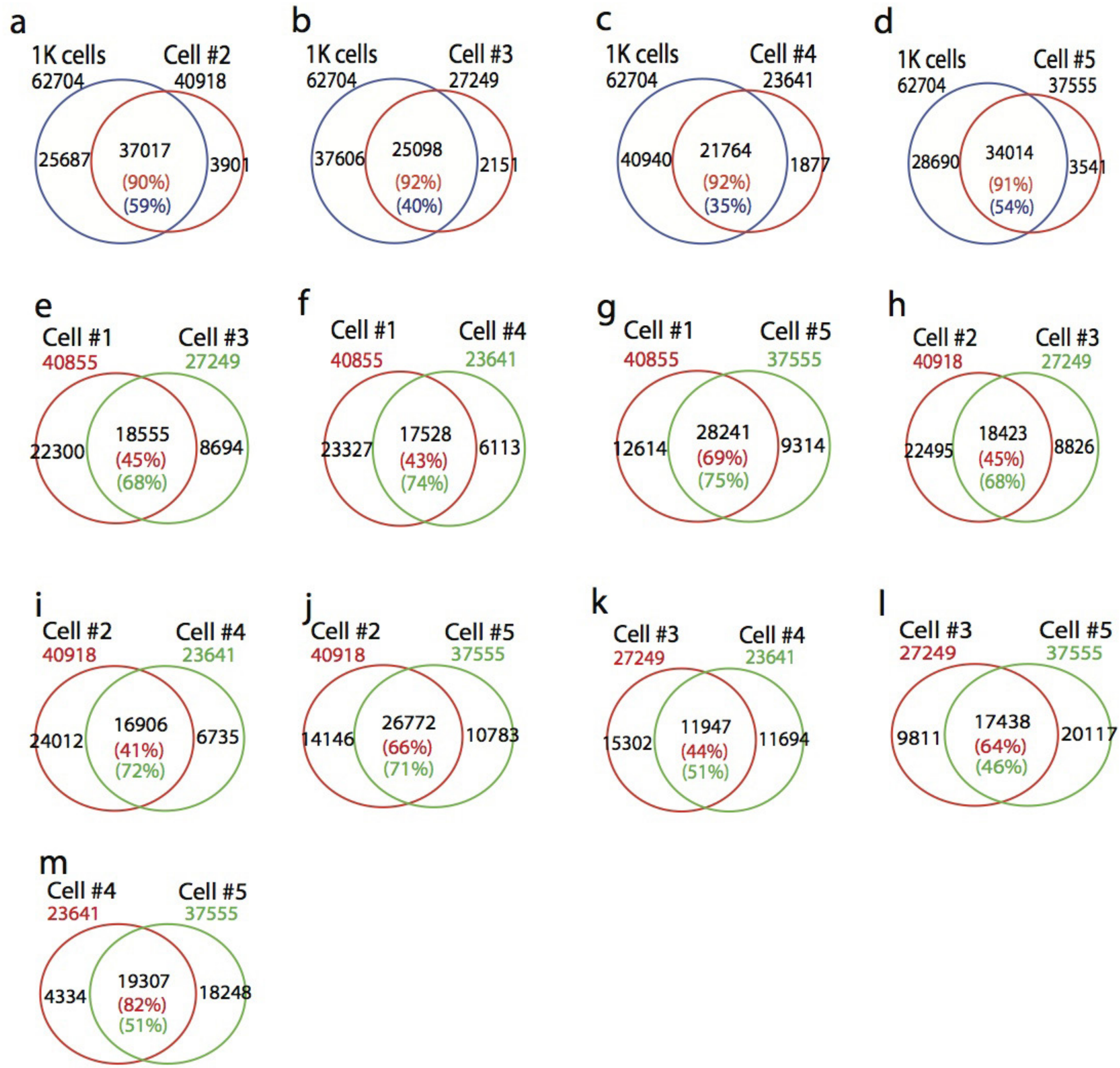
Identifying tumour-specific mutation: we generated scDNase-seq libraries using tumour or their neighbouring cells recovered from FFPE tissue section slides. The sequence reads were mapped by Bowtie2 (ref. 32). to the human reference genome (hg18). The paired reads with distance $<$ 500 bp were kept if paired-end sequencing was performed. Then reads with MAPQ $<$ 20 and possible duplication were removed by SAMtools[39]. Variation calling on each normal–tumour pair was conducted using SAMtools mpileup, with diploid model, MAPQ $\geq$ 20 and base alignment quality (BAQ) $\geq$ 30. The variations that only normal and tumour show different genotypes were kept. Then the low-quality variations were filtered (query quality (QUAL) $<$ 20, mapping quality (MQ) $<$ 20, phred probability of all samples being the same (FQ) $<$ 0, variant distance bias (VDB $<$ 0.01 and minor allele $<$ 3). We obtained 31 variation candidates in FTC 440 (Supplementary Table 13), many of them located on the predicted transcription-factor binding motifs.

Tumour- and normal-cell-specific DHS: the genome-wide DHSs were obtained by peak calling of the normal cell and tumour cell scDNase-seq libraries, respectively. The DHSs in normal cells and tumour cells were pooled, and reads in each library among the pooled DHSs were counted. The normal- and tumour-cell-specific DHSs were identified using EdgeR.

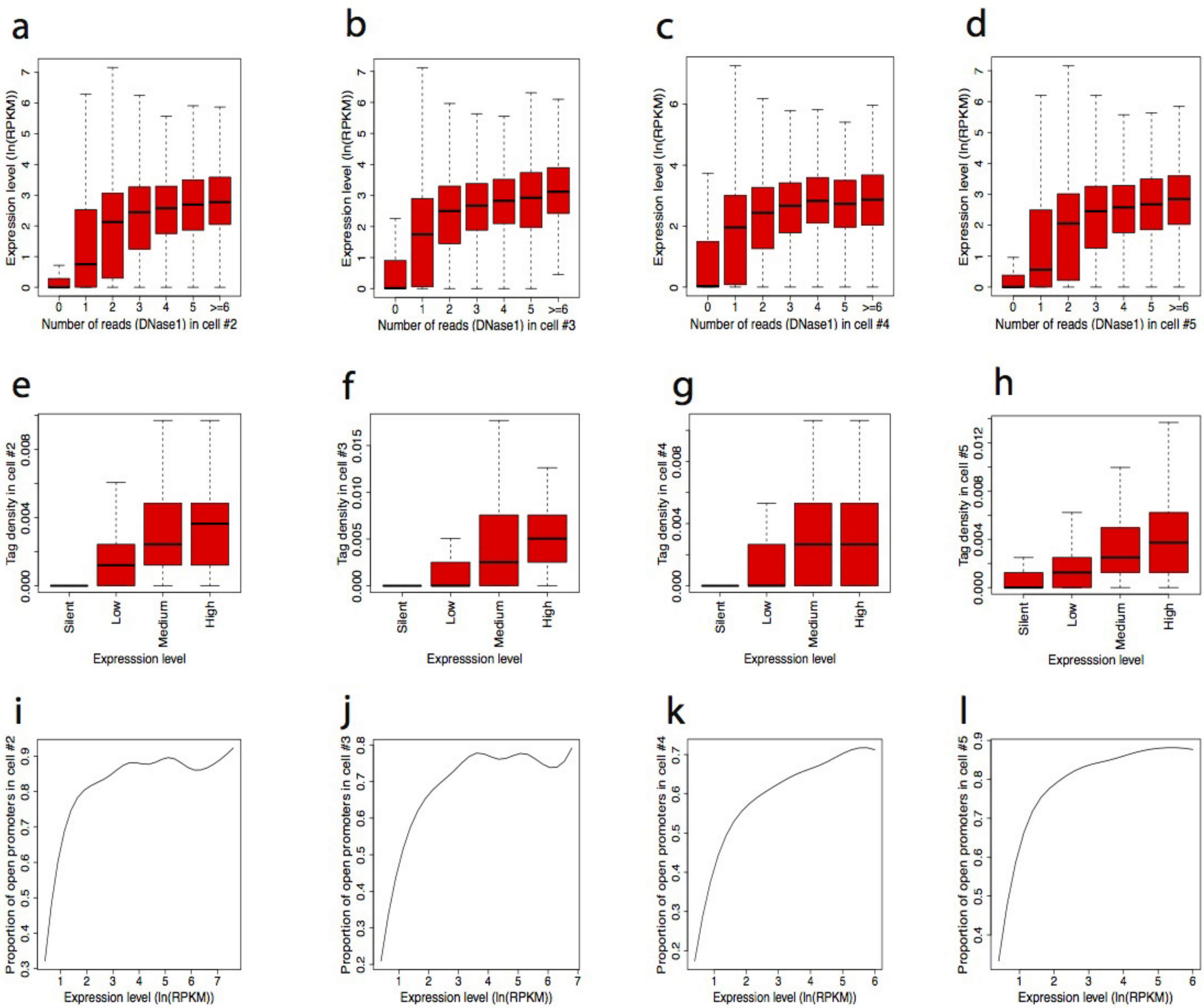
28. Hu, G. *et al.* H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* **12,** 180–192 (2013).
29. Hu, G. *et al.* Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nature Immunol.* **14,** 1190–1198 (2013).
30. Kraushaar, D. C. *et al.* Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. *Genome Biol.* **14,** R121 (2013).
31. Stamatoyannopoulos, J. A. *et al.* Mouse ENCODE Consortium. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol.* **13,** 418 (2012).
32. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).
33. Zhang, Y. *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9,** R137 (2008).
34. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
35. Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21,** 650–659 (2005).
36. Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-seq data. *Bioinformatics* **25,** 1952–1958 (2009).
37. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
38. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnol.* **28,** 495–501 (2010).
39. Li, H. *et al.* 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).



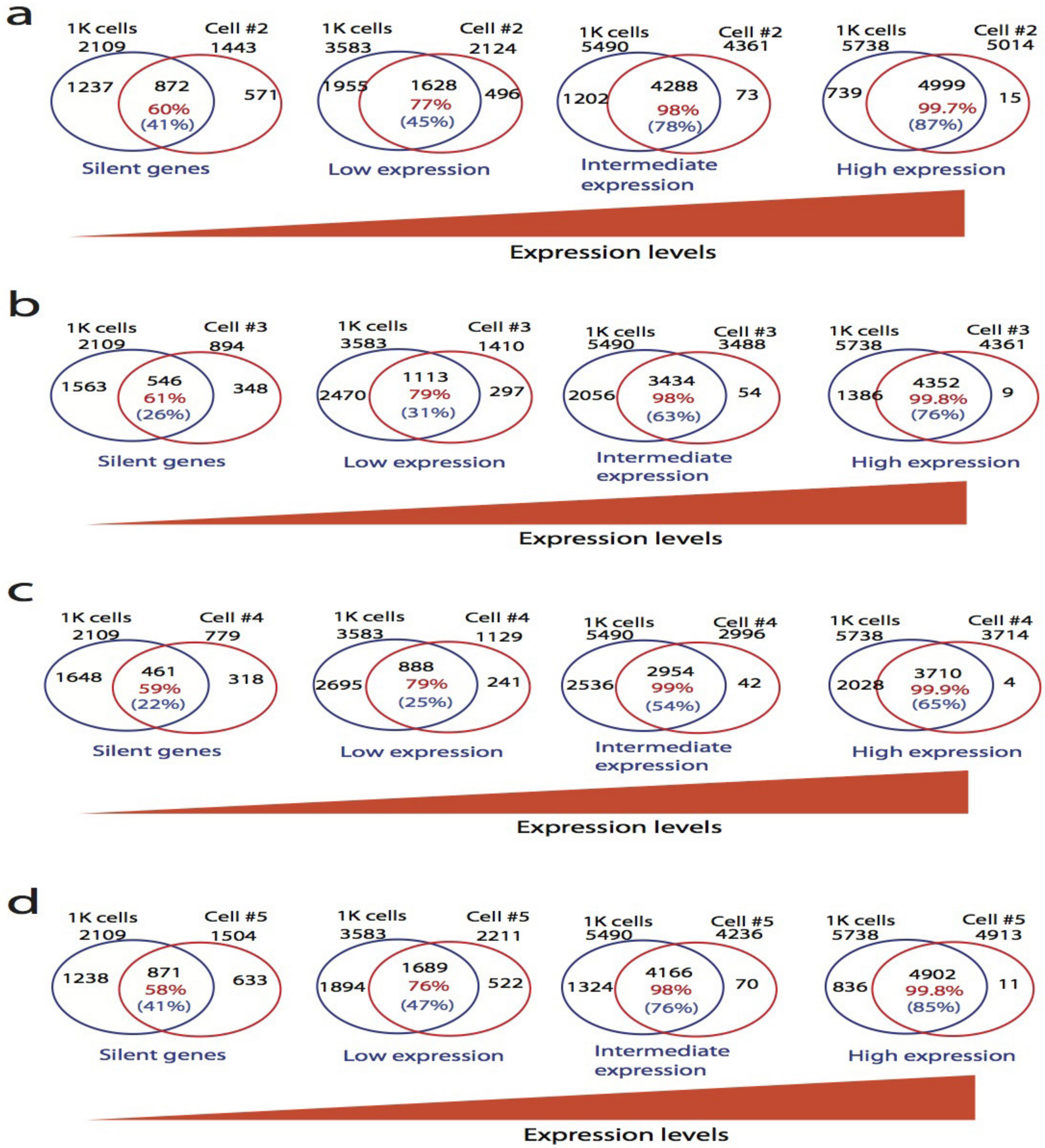
Extended Data Figure 1 | Scatter plots showing the high correlation of DHSs detected in a small number of cells or single cells. Each dot represents the tag density of one DHS or more DHSs with the same value. **a–c**, Correlation of DHSs between ENCODE data and small cell number libraries. **d–l**, Correlation of DHSs between the five single cells.



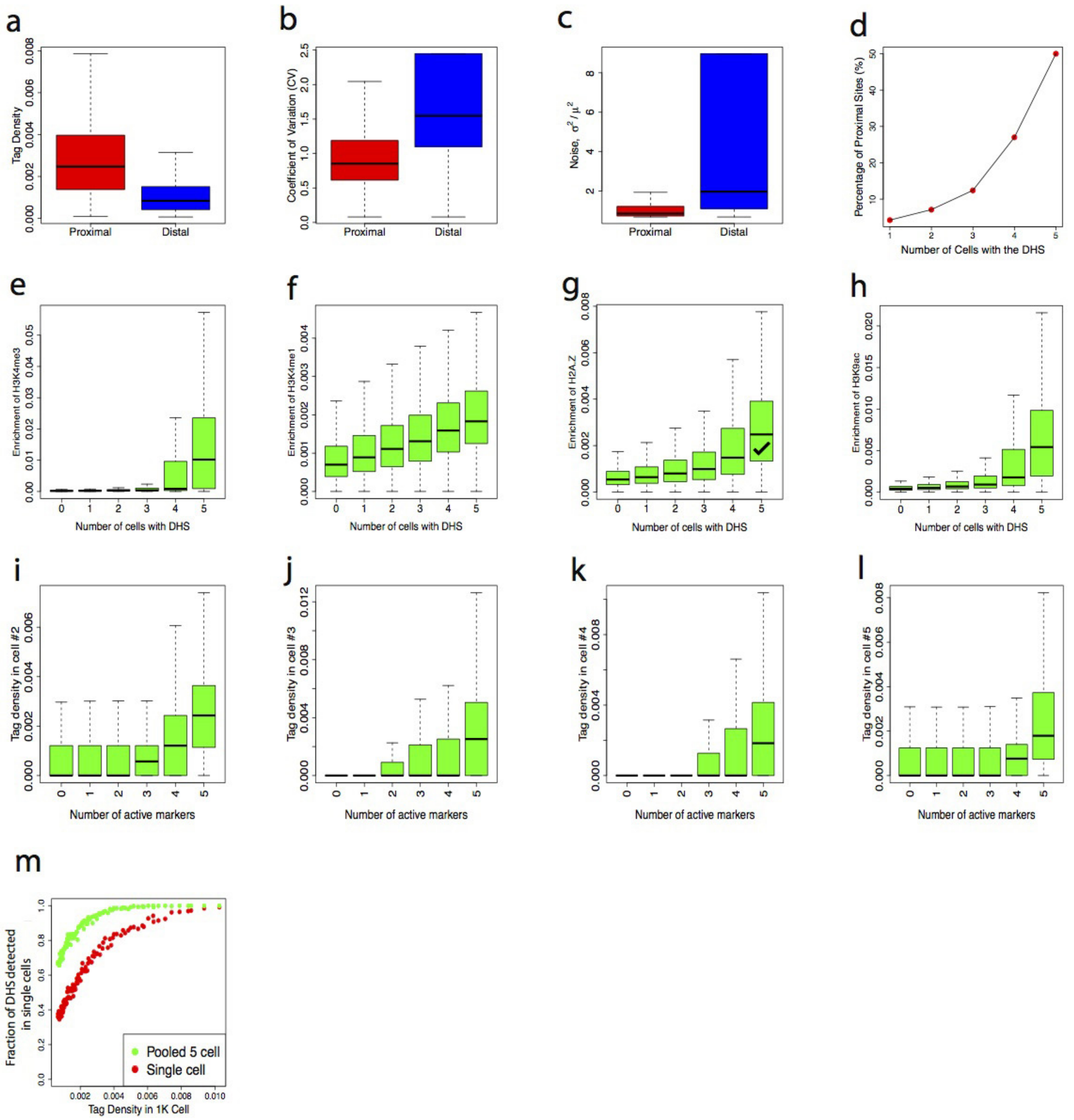
Extended Data Figure 2 | Venn diagrams showing DHSs detected in single cells significantly overlap with those detected in 1,000 cells or the other single cells. The total number of DHSs in each library is indicated outside the Venn diagrams. **a–d**, Overlapping DHSs between single-cell and 1,000-cells data. **e–m**, Overlapping DHSs between two single cells.



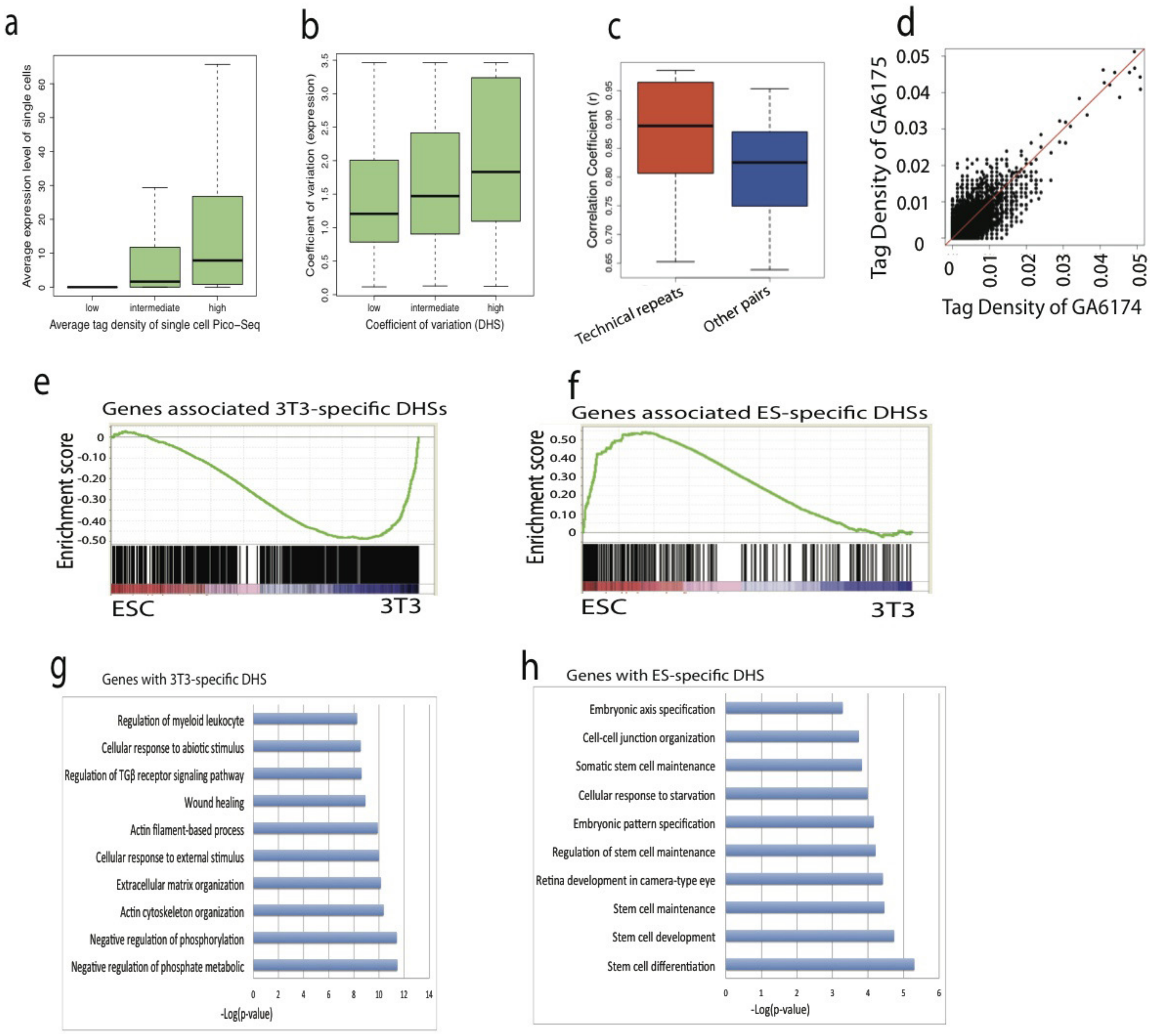
Extended Data Figure 3 | Detectability of single-cell DHSs at promoters is positively correlated with gene expression. a–d, Detection of DHSs around TSSs in each single cell is correlated with higher gene expression. Genes were sorted according to the number of scDNase-seq reads within a ±1 kb region of TSSs and plotted against their expression on the *y* axis. **e–h**, scDNase-seq tag density around TSSs in each single cell is positively correlated with gene expression. Genes were sorted into four groups according to their expression levels. Box plots show scDNase-seq tag density around TSSs (*y* axis). **i–l**, The proportion of open promoters detected by scDNase-seq in each single cell is positively correlated with gene expression.



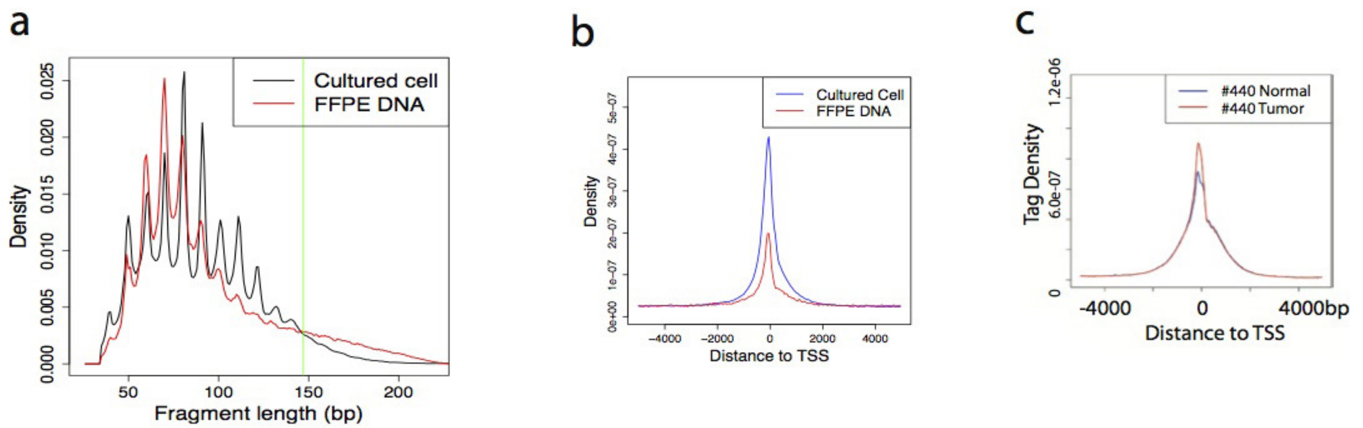
Extended Data Figure 4 | The DHSs on the promoter of highly expressed genes are more reproducible. The percentage of overlap between DHSs detected in 1,000 cells and each single cell positively correlates with gene expression. The total number of silent genes, lowly expressed, intermediately expressed and highly expressed genes showing DHSs are indicated outside the Venn diagrams. The numbers in red indicate the percentages of DHSs detected in a single cell that overlapped with the DHSs detected in 1,000 cells.



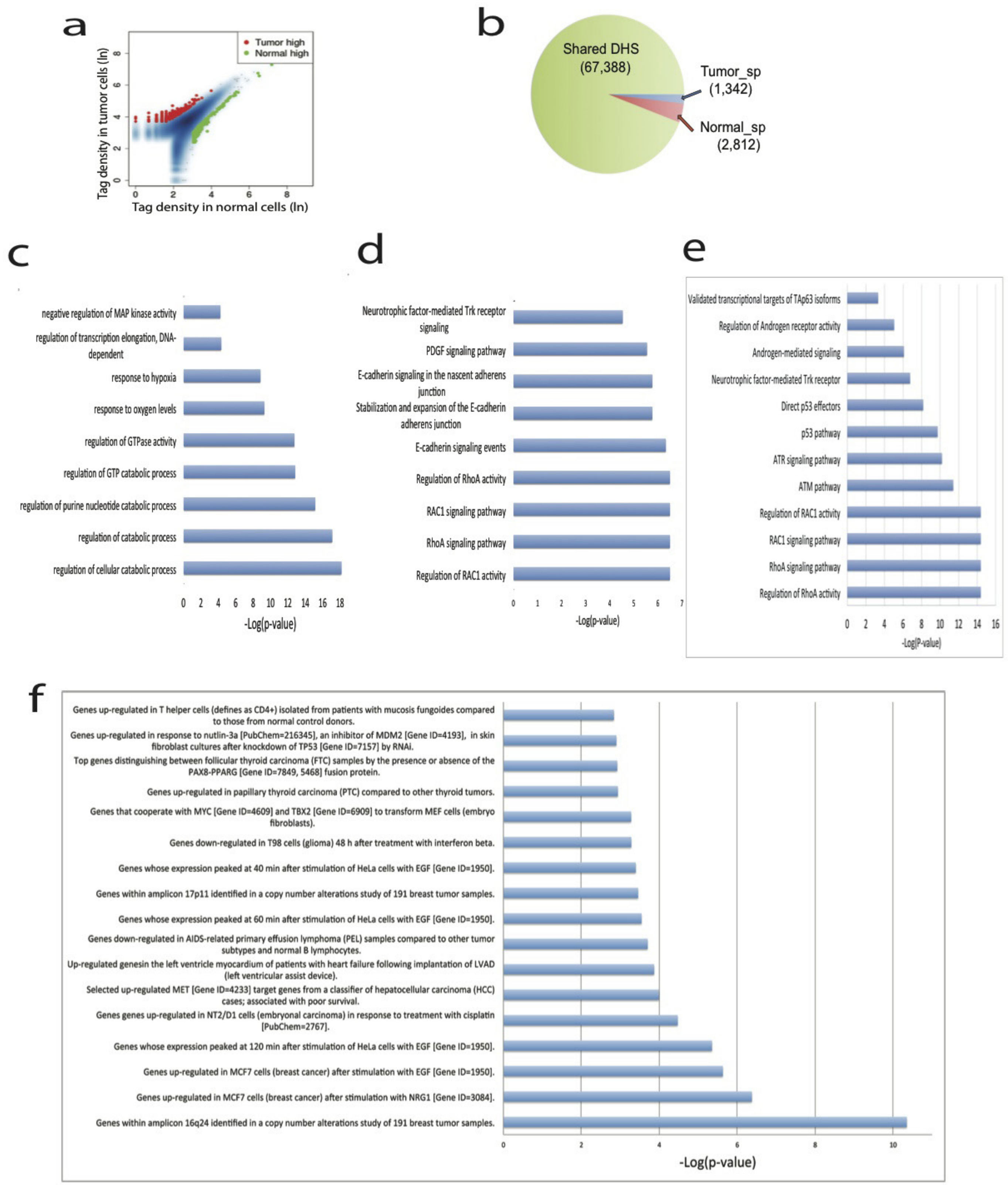
Extended Data Figure 5 | The high variations of distal DHSs and detectability of single-cell DHSs correlate with the number of active histone modifications. a–d, Distal DHSs showing lower tag density and higher variation than that of proximal DHSs among single cells. The average tag densities of the proximal DHSs among single cells are higher than those of distal DHSs (**a**). The proximal DHSs showed much higher variation (**b**) and noise (**c**) than those of distal DHSs. The fraction of proximal DHSs highly correlates with the number of cells with the DHSs (**d**). **e–h**, The histone modification levels (H3K4me3, H3K4me1, H3K9ac and H2A.Z) correlate with the detectability of DHSs in single cells. Histone modification peaks were sorted according to the number of single cells in which they were detected by scDNase-seq. The active histone modification enrichment levels of each group are displayed using box plots. **i–l**, scDNase-seq density in each NIH3T3 single cell positively correlates with the number of active histone modifications at the DHS. Tag density of each scDNase-seq was sorted according to the number of histone modifications measured on a population of cells using ChIP-seq. The scDNase-seq tag densities for each group are shown by box plots. **m**, The DHS detectability in single cells is correlated with the tag density of DHS peaks in the 1,000-cells library. The DHSs obtained from the 1,000-cells library were binned to 100 groups on the basis of the tag density (or peak height) (*x* axis). The *y* axis indicates the fraction of DHSs detected in single-cell or pooled five single-cell scDNase-seq libraries for each bin.



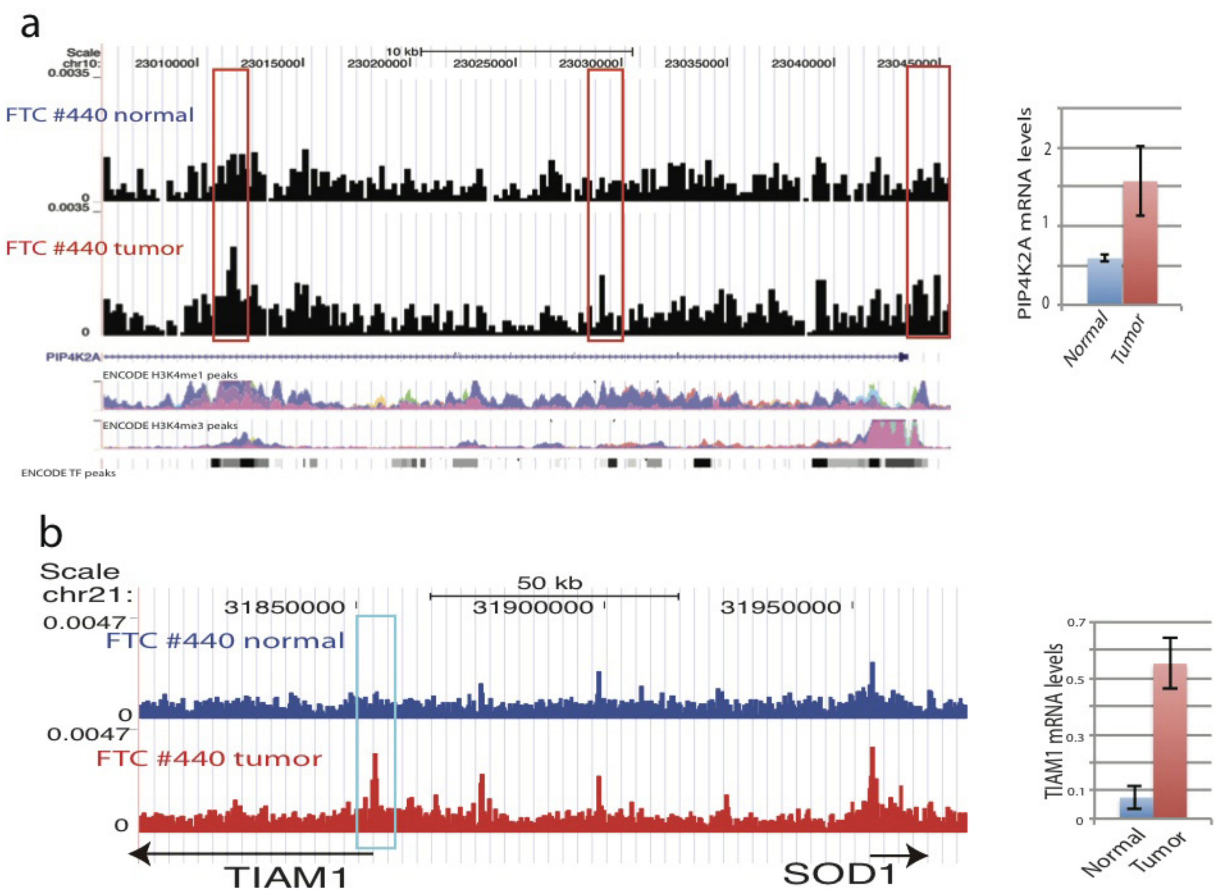
Extended Data Figure 6 | Biological variations contributed to cell-to-cell variation and DHSs detected in single-cell scDNase-seq can predict cell-type-specific enhancers. a, b, Single-cell chromatin accessibilities and single-cell gene expressions are positively correlated. Average (**a**) and variation (**b**) of tag density in single-cell scDNase-seq at gene promoters correlates with that of gene expression level in single cells, respectively. **c**, **d**, Biological variations contribute to cell-to-cell variations because the correlation coefficients between technical repeats are significantly higher than those of other pairs of libraries (non-technical repeat pairs). Two NIH3T3 cells were sorted into one tube, which were digested with DNase I and then split to two tubes. Thus each tube contained the amount of DNA that should have been similar to that of one cell. By doing this, the two libraries prepared using the two tubes could be treated as technical repeats. **c**, Correlation coefficients between technical repeats are higher than those of other pairs of libraries. **d**, Scatter plot of a pair of technical repeats. **e**, **f**, Genes associated with NIH3T3- and ESC-specific DHSs correlate with NIH3T3- and ESC-specific expressed genes, respectively. **g**, **h**, Genes associated with NIH3T3- or ESC-specific DHSs are enriched in distinctive GO terms.



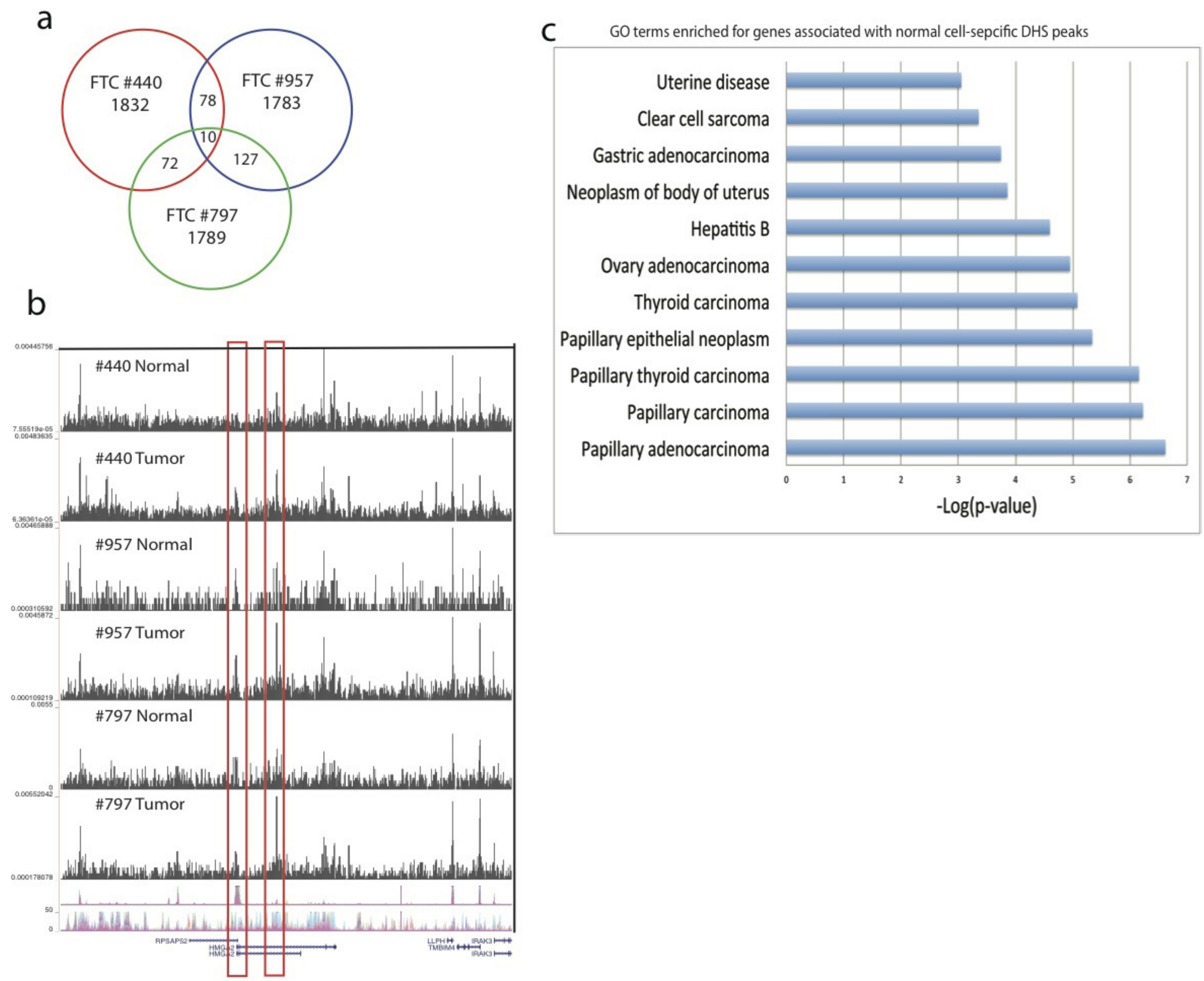
Extended Data Figure 7 | The scDNase-seq libraries for FFPE tissue slides showing expected patterns. **a**, DNA fragments of the scDNase-seq libraries both from cultured cells and from FFPE tissues show periodical cut patterns expected from DNase I digestion of nucleosomal DNA. **b**, The scDNase-seq reads both from cultured cells and from FFPE tissues are enriched around TSSs. **c**, The read enrichments around TSSs of FTC 440 normal and FTC 440 tumour are similar.



Extended Data Figure 8 | Tumour-specific DHS in FTC-440-enriched GO terms and pathways. **a**, **b**, Normal- and tumour-specific DHSs account for a small fraction of the total DHSs. **c**, GO biological process terms are significantly enriched in the tumour-specific DHS. **d**, Pathways are significantly enriched in the tumour-specific DHS. **e**, Pathways are significantly enriched in the tumour-specific DHS with relaxed threshold. **f**, Gene sets that represent gene expression signatures of genetic and chemical perturbations are significantly enriched in the tumour-specific DHS.



Extended Data Figure 9 | Tumour-specific DHSs are correlated with increased expression in the tumour cells. a, Genome Browser image showing the increased chromatin accessibility of two tumour-specific DHS regions at the *PIP4K2A* gene locus (left). The ENCODE H3K4me1 and H3K4me3 peaks are shown at the bottom of the panel. The PIP4K2A mRNA levels in normal and tumour cells, respectively, determined by qRT–PCR and normalized to GAPDH (right). **b**, Genome Browser image showing the increased chromatin accessibility of the TIAM1 promoter in thyroid tumour cells (left). TIAM1 mRNA levels in normal and tumour cells, respectively, determined by qRT–PCR and normalized to GAPDH (right).



Extended Data Figure 10 | Tumour-specific DHSs in each individual are usually unique. a, The vast majority of DHSs are unique to each individual tumour case. **b**, Genome Browser image showing the two tumour-specific DHSs at the *HMGA2* locus in three patients with FTC. **c**, The normal cell-specific DHSs are enriched in multiple disease ontologies in PTC 131.

LETTER

doi:10.1038/nature15545

Transcriptional regulators form diverse groups with context-dependent regulatory functions

Gerald Stampfel[1], Tomáš Kazmar[1], Olga Frank[1]†, Sebastian Wienerroither[1], Franziska Reiter[1] & Alexander Stark[1]

One of the most important questions in biology is how transcription factors (TFs) and cofactors control enhancer function and thus gene expression. Enhancer activation usually requires combinations of several TFs[1], indicating that TFs function synergistically and combinatorially[2,3]. However, while TF binding has been extensively studied, little is known about how combinations of TFs and cofactors control enhancer function once they are bound. It is typically unclear which TFs participate in combinatorial enhancer activation, whether different TFs form functionally distinct groups, or if certain TFs might substitute for each other in defined enhancer contexts. Here we assess the potential regulatory contributions of TFs and cofactors to combinatorial enhancer control with enhancer complementation assays. We recruited GAL4-DNA-binding-domain fusions of 812 *Drosophila* TFs and cofactors to 24 enhancer contexts and measured enhancer activities by 82,752 luciferase assays in S2 cells. Most factors were functional in at least one context, yet their contributions differed between contexts and varied from repression to activation (up to 289-fold) for individual factors. Based on functional similarities across contexts, we define 15 groups of TFs that differ in developmental functions and protein sequence features. Similar TFs can substitute for each other, enabling enhancer re-engineering by exchanging TF motifs, and TF–cofactor pairs cooperate during enhancer control and interact physically. Overall, we show that activators and repressors can have diverse regulatory functions that typically depend on the enhancer context. The systematic functional characterization of TFs and cofactors should further our understanding of combinatorial enhancer control and gene regulation.

We sought to characterize the potential regulatory contributions of different TFs to combinatorial enhancer control, that is, the regulatory functions of the TF proteins following DNA binding, regardless of their specific roles *in vivo*. We reasoned that such contributions could be best assessed using ectopic tethering assays in the context of DNA sequences that closely resemble active enhancers, ideally only lacking the input of a single TF. In particular, such a setup may allow the assessment of obligate combinatorial factors whose regulatory activities depend on partners and which would otherwise appear non-functional. We therefore developed enhancer complementation assays based on activator bypass experiments[4] used to test candidate TF or cofactor function in transcription control and to dissect promoters[5–10].

We mutated TF-binding-motif sequences within active enhancers to 'upstream activating sequence' (UAS) motifs for the GAL4 DNA-binding domain (GAL4-DBD), recruited 474 *Drosophila* TFs[11] via GAL4-DBD fusion proteins to the positions of the mutated motifs (enhancer context), and measured enhancer activities by luciferase assays in S2 cells, normalizing to GFP recruitment (Fig. 1a). Since expression and recruitment is standardized, the factors' regulatory functions can be assessed in a highly controllable manner independently of the factors' endogenous expression and DNA binding. Overall, the assays were highly reproducible: 75% of all data points had standard deviations (s.d.) $<10\%$, and 95% of the s.d. were $<18\%$ across four biological replicates.

We started with an enhancer that was highly active in *Drosophila* S2 cells and for which mutations of CGCG- or GATA-type motifs strongly reduced its activity[3] (Fig. 1b). We replaced either the CGCG- or the GATA-motifs with UAS motifs (Fig. 1a) and assessed which TFs restored enhancer function or repressed the remaining basal activities. Of the 474 TFs, 100 were activating ($\geq$1.5-fold compared to GFP; $P < 0.05$ false discovery rate (FDR)-corrected for 474 tests) in the CGCG- and 84 in the GATA-context (Fig. 1c), including TFs that recognize the CGCG- and GATA-motifs, respectively (Extended Data Table 1). This compares to 77 TFs that activated on their own (that is, when recruited to UAS motifs outside an enhancer context), suggesting that TF function might be context-dependent. Indeed, 46 TFs activated the CGCG context at least 1.5-fold ($P < 0.05$) more strongly than the GATA context, even though both contexts were derived from the same enhancer (Fig. 1d and Extended Data Fig. 1). To test if native untagged TFs recapitulate these results, we chose Ets at 21C (Ets21C), Deformed (Dfd), and Hairy/E(spl)-related with YRPW motif (Hey) that preferentially activated the CGCG context to different extents (Fig. 1d). We replaced the CGCG- and GATA-motifs with binding sites for these TFs and expressed the untagged TFs (Fig. 1e). This activated the mutant enhancers in a manner consistent with the results from GAL4-DBD-mediated recruitment: Ets21C and Dfd activated only the CGCG context, while Hey activated both (CGCG 1.3-fold more highly), confirming the similarity and context-dependency of these TFs' regulatory functions.

Intrigued by the context-dependency of some TFs even within a single enhancer, we decided to include more diverse regulatory contexts (Extended Data Fig. 1). We created 19 motif-mutant enhancer contexts for different types of TF motifs and different enhancers with broad, cell-type-specific[3], or hormone-inducible[12] activities. We also added five contexts consisting of UAS sites and core promoters specific towards developmental or housekeeping enhancers, respectively[13].

Nearly half of all TFs (42%) were activating and most (93%) of the remaining 276 TFs were repressing in at least one of the 24 contexts ($\geq$1.5-fold; $P < 0.05$ FDR-corrected for 24×474 tests), suggesting that most TF-fusion proteins were functional. Many TFs had similar regulatory effects across the 24 contexts, suggesting that they might be functionally equivalent. We grouped all TFs into 15 clusters using unsupervised spectral clustering (Fig. 2a and Supplementary Table 1) and confirmed that these clusters are robust to bootstrapping and reproducible when using independent biological replicates (Extended Data Fig. 2). This revealed clusters of diverse regulatory functions (Fig. 2b and Extended Data Fig. 3), including cluster 8 with TFs that activated in most contexts (global activators) such as Antennapedia, Sox14 and Sox15, Clock (Clk), and Zelda, and clusters 3, 5 and 7 with global repressors (for example, Snail, Runt, Engrailed and Kruppel). These TFs seemed to dominate or override other regulatory cues, consistent with their ability to function in isolation. TFs of other clusters were only

[1]Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Dr. Bohr-Gasse 7, 1030 Vienna, Austria. †Present address: Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstraße 108, 01307 Dresden, Germany.

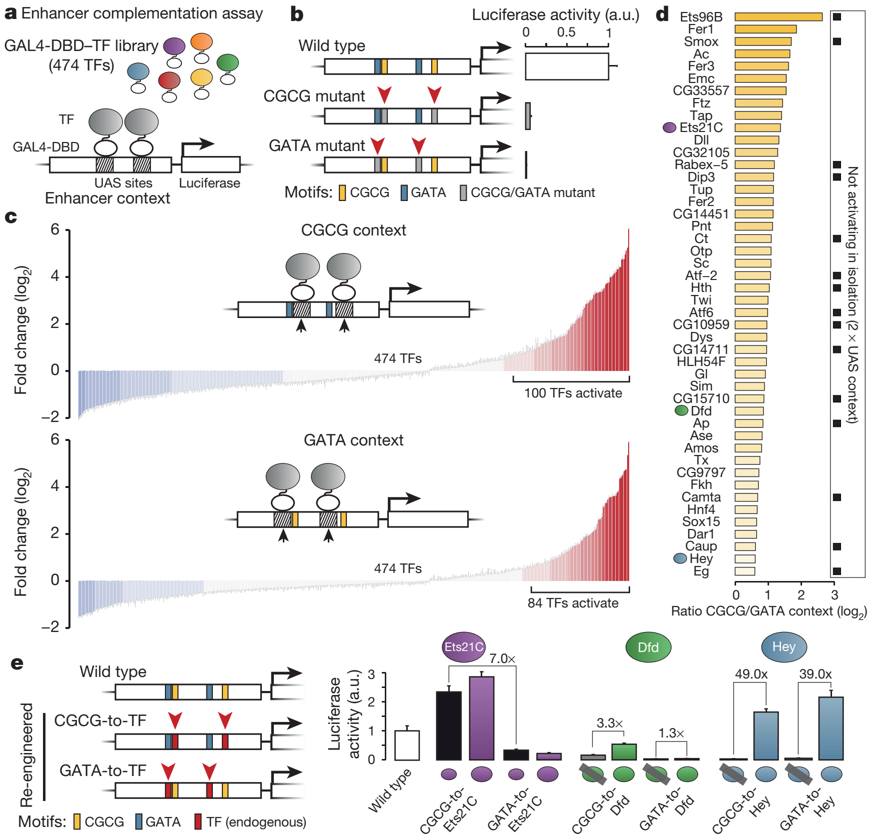


Figure 1 | Enhancer complementation assays for 474 TFs. a, Schematic overview. **b**, Activity of an enhancer and motif-mutant variants (data from ref. 3). a.u., arbitrary units. **c**, Enhancer complementation assays for CGCG- and GATA-contexts (normalized luciferase values for 474 TFs; red: activation, blue: repression). **d**, Preferential activation of CGCG- versus GATA-context (≥1.5-fold, FDR-corrected $P < 0.05$). Black boxes, TFs inactive on their own; colour, TFs tested in **e**. **e**, Validation of context-dependent TFs (details see text). Luciferase activities (firefly/*Renilla*) with (colours) or without (grey and black (*Ets21C* is expressed in S2 cells)) co-transfection of the respective untagged TF. Error bars show standard deviations ($n = 4$, biological replicates).

weakly active in the contexts tested, and, notably, the TFs of some clusters were context-dependent. For example, cluster 10 TFs preferentially activated the housekeeping core promoter and might constitute factors of a distinct transcriptional program[13], including Myb-interacting protein 120 (Mip120) and CG6813, the cluster's strongest activator (21-fold; $P = 4.4 \times 10^{-4}$). In contrast, cluster 1 TFs preferentially activated hormone-receptor contexts, that is, when recruited to ecdysone receptor (EcR)-binding sites in enhancers inducible by the insect steroid hormone ecdysone. Examples are Twist, Reversed polarity, Pointed, and other developmental TFs.

Intrigued by TFs that preferentially activated hormone-receptor contexts, we selected four such TFs from clusters 1 and 15, Ets96B, Helix loop helix protein 4C (HLH4C), Atonal (Ato), and Glass (Gl), and asked whether replacing the EcR motif with the motifs of these TFs would activate the enhancer in a TF-dependent but hormone-independent manner. This was indeed the case for all four TFs, and the effect was specific to the combination of motif and enhancer context (Fig. 2c), suggesting that these TFs might contribute regulatory functions equivalent to the activated EcR.

To assess the TF clusters independently of our approach, we asked whether they were enriched in Gene Ontology (GO) categories or in protein sequence features such as Pfam domains or short peptide motifs. Indeed, many such features were differentially distributed between the clusters ($P < 0.01$; empirical FDR = 0.1) and each cluster was enriched for at least one such feature (Fig. 2d and Extended Data Fig. 4; Supplementary Table 2). As expected, amino acid repeats known to mediate activation (for example, poly-glutamine[14]) or repression (for example, poly-alanine[15]) were enriched in activating clusters (1, 8) and repressing clusters (3, 5, 7), respectively. However, of several activating clusters, only cluster 1 was enriched in GO categories relating to development, suggesting a preferential use of cluster-1-type TFs during developmental gene regulation, which presumably relates to these TFs' dependence on partner TFs, enabling combinatorial control. Similarly, only repressing cluster 7 but not 3 or 13 was enriched in GO categories relating to Notch signalling, cardiocyte differentiation, or morphogenesis, suggesting that repression might occur through various means that are differently employed *in vivo*. Indeed, the three repressing clusters also differed in the enrichment of peptide motifs known to bind the co-repressors C-terminal binding protein (CtBP; cluster 7) or Sin3A (cluster 3), suggesting a functional association between the TFs in these clusters and the respective co-repressors (see below).

These results show that the different TF clusters, obtained solely based on the TFs' context-dependent regulatory functions, differ in several other aspects, which lends independent support to the clustering. It also suggests that the respective TFs are differentially employed *in vivo* (for example, during development), and that their distinct functions might arise through the recruitment of different types of cofactors (for example, CtBP versus Sin3A).

To assess the regulatory activities and the clustering in different cell types, we tested 171 TFs (9 to 17 TFs from each cluster) across six contexts in Kc167, BG3 and ovarian somatic cells derived from embryos, larvae and adult ovaries, respectively. These cell types differ increasingly from S2 cells in gene expression, enhancer activities, and the enhancers' motif signatures[3], yet TF activities were remarkably similar: all 18 pairwise comparisons had Pearson correlation coefficients (PCCs) ≥0.5 and 15 had PCCs ≥0.8 (all $P < 1 \times 10^{-3}$; Extended Data Fig. 5).

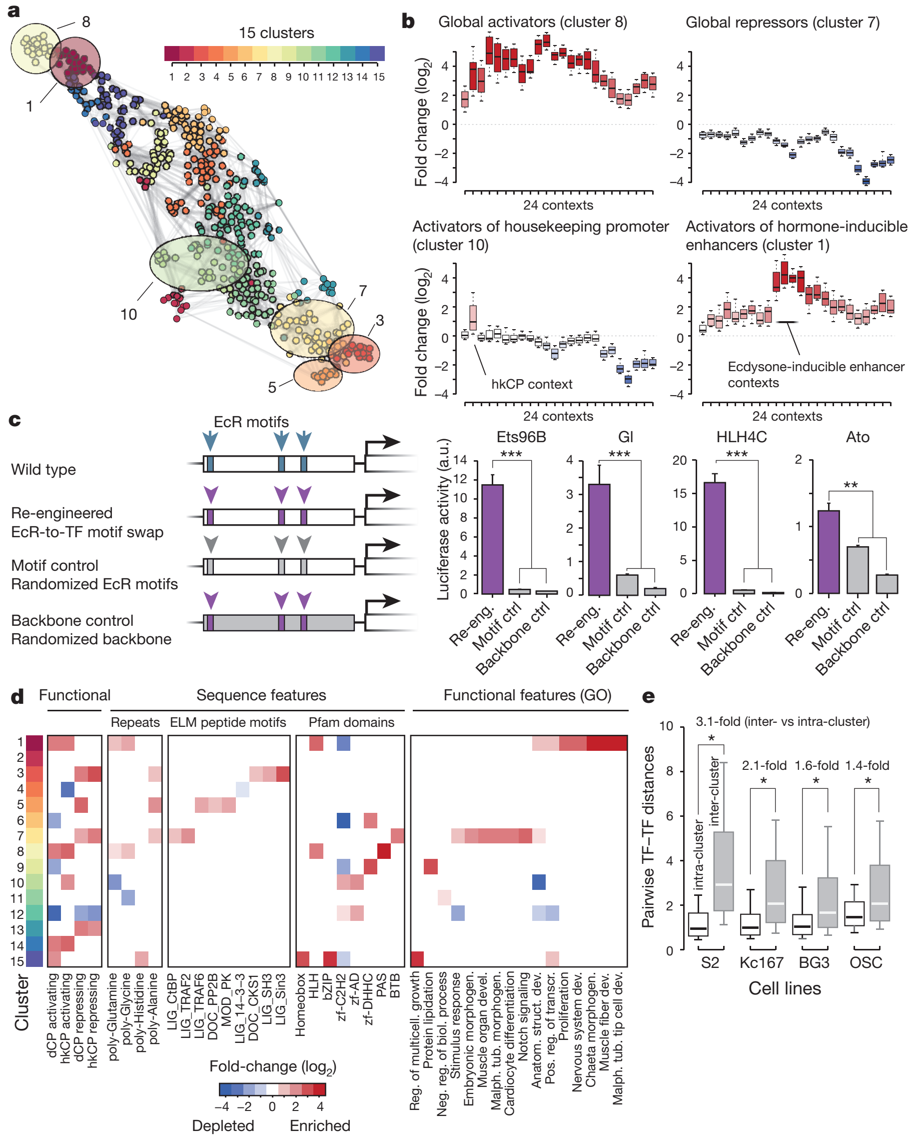


Figure 2 | TFs have diverse regulatory functions. a, 15 TF clusters (see text for highlighted clusters). **b**, Normalized luciferase values across 24 contexts (selected from Extended Data Fig. 3). **c**, Validation of hormone-context-preferential TFs (luciferase activities (firefly/*Renilla*) for re-engineered enhancers (re-eng.; purple) and controls (grey; see schematic)). Error bars show standard deviations ($n = 4$, biological replicates). $^{**}P < 1 \times 10^{-2}$; $^{***}P < 1 \times 10^{-3}$. **d**, Enrichments for TFs that activate or repress the 4×UAS-dCP or 4×UAS-hkCP contexts ('Functional'), protein-sequence features and GO-categories; significant ($P < 0.01$; empirical FDR = 0.1) enrichments, red; and depletions, blue; others, white; see Extended Data Fig. 4 and Supplementary Table 2 for details and all data. **e**, Pairwise distances between activity profiles in Kc167, BG3 and OSC cells support functional TF clusters (all empirical $P < 10^{-6}$, indicated with single asterisks).

Moreover, the TFs' activity profiles across the 6 contexts support the original clustering in each cell type: pairs of TFs from within original clusters were more similar than pairs across clusters (≥1.4-fold and $P < 1 \times 10^{-6}$ for all four cell lines; Fig. 2e). Intrigued by these results, we tested the ability of 107 *Drosophila* TFs (90) and cofactors (17) to activate transcription in human HeLa cells and found a good quantitative agreement across all factors (PCC = 0.74; Extended Data Fig. 6). These results suggest that many TFs function predominantly, but not completely, independently of cell type. We note that alternative splicing and post-translational modifications (for example, downstream of cellular signalling pathways) probably alter and diversify the regulatory functions of individual TFs.

The regulatory functions of TFs are generally mediated through transcriptional cofactors, which typically lack DNA-binding domains and are recruited to enhancers by TFs. To assess whether cofactor functions are similarly diverse or potentially more uniform, we cloned 338 putative cofactors from diverse protein families (Fig. 3a and Supplementary Table 3) and tested their activating and repressing functions in all 24 contexts using GAL4-DBD-mediated recruitment as for TFs (Fig. 3b).

Most cofactors (80%) were sufficient to activate or repress transcription in at least one context (≥1.5-fold; $P < 0.05$ after FDR correction for 24 × 338 tests), and the activities of well-studied factors matched their known functions (Fig. 3c–e): for example, P300 (also known as Nejire) strongly activated transcription in all contexts, as did the histone-methyltransferase Lost PHDs of Trr (Lpt) of the Set1/COMPASS-like complex, and the Mediator subunits MED15 and MED25, while the co-repressors CtBP, Sin3A and CoRest were strongly repressing in all contexts. Other cofactors had context-specific functions, including Chromator (Chro), TBP-associated factor 4 (Taf4) and Trithorax-related (Trr), which preferentially activated the housekeeping core promoter (Trr was even repressing in all other contexts). Chro is part of the non-specific lethal (NSL) complex which activates genes involved in cell proliferation and DNA replication[16] and Taf4 is important at TATA-less promoters[17]. Similar to the corresponding TFs above, these cofactors might be part of a dedicated housekeeping regulatory program[13].

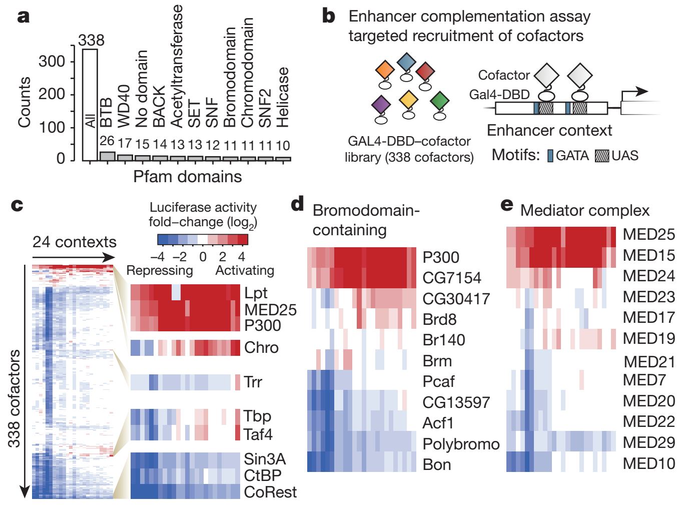


Figure 3 | Transcriptional cofactors can be sufficient for activation and repression and have context-dependent regulatory functions. **a**, Pfam-domain content of the 338 putative cofactors tested. **b**, Scheme depicting enhancer complementation assays for cofactors. **c**, Overview of the regulatory activities for cofactors across all 24 contexts (bi-clustered 338 × 24 heat map) and zoom for selected cofactors. **d**, **e**, Diverse regulatory activities for bromodomain-containing cofactors (**d**) and subunits of the Mediator complex (**e**) (see Extended Data Fig. 7 for additional complexes).

Notably, different members of a single complex or domain family frequently had different regulatory functions (Fig. 3d, e and Extended Data Fig. 7), cautioning against the transfer of annotations based on these grounds. For example, the activities of bromodomain-containing (BRD) cofactors (Fig. 3d) range from the strongly activating P300 and CG7154, the orthologue of human BRD7 and BRD9, to the context-dependent CG30417, and to Polybromo that was strongly repressing in most contexts, consistent with its previous implication in transcriptional repression[18]. Similarly, Mediator subunits MED15 and MED25 were strongly activating in most contexts, MED23 and MED24 were context-dependent, and MED29 was repressing, consistent with the function of human MED29[19] (Fig. 3e). This suggests that different TFs might interact with the Mediator complex through distinct subunits, or that complexes with variable composition and function might exist. Consistently, MED15 and MED25 interact directly with strong activators (for example, GAL4 and VP16[20,21]) and MED23 and MED24 are involved in signal-dependent and hormone-induced transcription, respectively[22,23].

The activating and repressing effects across 24 contexts were highly similar for many TFs and cofactors (Fig. 4a), which provided a means to infer functional associations (Fig. 4b). As expected, P300, Lpt, MED15 and MED25 were assigned to globally activating TFs (cluster 8) and context-dependent cofactors to context-dependent TFs (for example, Mip120 and Bsh to cluster 10 and Chro to cluster 14). Interestingly however, the globally repressing cofactors Sin3A and CtBP were assigned to different clusters of repressing TFs (cluster 3 versus 7), in agreement with the differential enrichment of peptide motifs involved in Sin3A and CtBP recruitment (Fig. 2d). Many of the assignments are consistent with known physical interactions, including the interaction between Chro and Pzg[24,25] or CtBP and Sna[26]. Indeed, the assignments were enriched for interactions reported in large-scale studies that used yeast two-hybrid assays[27] or co-affinity purification[25,28] (between 1.4- and 3.0-fold; all $P \leq 0.05$; Fig. 4c). In addition, the human orthologues of TF–cofactor pairs interacted 1.8-fold ($P = 0.025$) more frequently than expected[29]. These results suggest that the TF–cofactor assignments reflect functional associations and predict that the enhancer activation obtained by TF recruitment

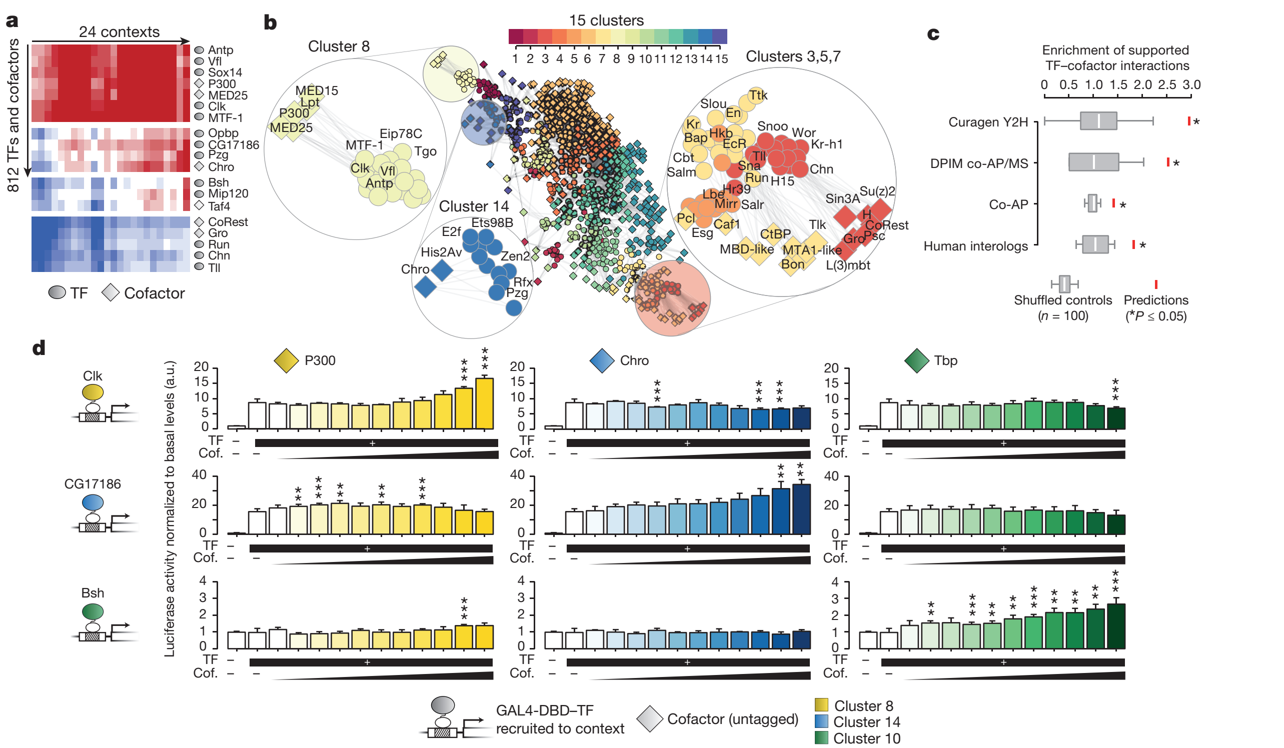


Figure 4 | TF–cofactor assignments through functional similarities. **a**, TFs (ovals) and cofactors (diamonds) show similar regulatory activities and localize next to one another in a bi-clustered 812 × 24 heat map (selection shown; see Supplementary Table 1 for all activities). Coloured as in Fig. 3c–e. **b**, TF clusters (as in Fig. 2a) and assigned cofactors. Highlighted are clusters 8, 14 and 3, 5, 7 and assigned cofactors. **c**, TF–cofactor assignments are enriched for physical interactions (red bars) compared to shuffled assignments, for which the box-plots indicate the 10th, 25th, 50th, 75th and 90th percentiles; *hypergeometric $P \leq 0.05$. **d**, Boosting of Clk (top), CG17186 (middle) and Bsh (bottom) induced enhancer activities by untagged P300 (left), Chro (centre) and Tbp (right). Error bars show standard deviations ($n = 4$, biological replicates). **$P < 1 \times 10^{-2}$; ***$P < 1 \times 10^{-3}$.

should be boosted by the assigned but not by unrelated cofactors, even if the cofactors are not tethered via GAL4-DBD[30] (Fig. 4d). Indeed, in this experimental setup[30], the activation by Clk-recruitment (cluster 8) was further boosted by increasing amounts of untagged P300, but not by Chro or Tbp. In contrast, Chro specifically boosted CG17186 (cluster 14) and Tbp specifically boosted Bsh (cluster 10), consistent with the respective assignments.

Enhancer complementation assays provide a unique annotation and categorization of TFs and cofactors based on their regulatory functions, independent of the factors' endogenous roles and complementary to previous classifications through genetics, sequence comparisons, or genomics (for example, ChIP-seq). For many factors, including 266 putative TFs and cofactors ('CG' genes; Extended Data Fig. 8), our work provides the first functional characterization. All data are available at http://factors.starklab.org.

The existence of equivalence groups among TFs and cofactors with diverse context-dependent functions, even amid activators and repressors, has profound implications for our understanding of transcriptional gene regulation: while some enhancers might be controlled predominantly by individual activators, others may rely on specific combinations of distinct regulatory functions that are complementary and each insufficient for activation. It is therefore conceivable that different types of enhancers are controlled through non-overlapping sets of TFs and cofactors, enabling separate transcriptional programs even within individual cells (for example, ref. 13). The approach and categorization presented here provide a framework to dissect the molecular and biochemical nature of these functions and the mechanisms by which cooperativity at enhancers is established and transcriptional activation of target core promoters is achieved. Understanding these mechanisms will be crucial at a time when enhancer function and its control by TFs and cofactors are becoming increasingly central to our understanding of gene regulation in development and disease and the focus of novel therapeutic strategies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 January; accepted 2 September 2015.
Published online 9 November 2015.

1. Arnone, M. I. & Davidson, E. H. The hardwiring of development: organization and function of genomic regulatory systems. *Development* **124,** 1851–1864 (1997).
2. Han, K., Levine, M. S. & Manley, J. L. Synergistic activation and repression of transcription by *Drosophila* homeobox proteins. *Cell* **56,** 573–583 (1989).
3. Yáñez-Cuna, J. O. *et al.* Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat motifs as general enhancer features. *Genome Res.* **24,** 1147–1156 (2014).
4. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386,** 569–577 (1997).
5. Cheng, J. X., Gandolfi, M. & Ptashne, M. Activation of the Gal1 gene of yeast by pairs of 'non-classical' activators. *Curr. Biol.* **14,** 1675–1679 (2004).
6. Keung, A. J., Bashor, C. J., Kiriakov, S., Collins, J. J. & Khalil, A. S. Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. *Cell* **158,** 110–120 (2014).
7. Yuasa, Y. *et al. Drosophila* homeodomain protein REPO controls glial differentiation by cooperating with ETS and BTB transcription factors. *Development* **130,** 2419–2428 (2003).
8. Papadopoulos, D. K. *et al.* Functional synthetic *Antennapedia* genes and the dual roles of YPWM motif and linker size in transcriptional activation and repression. *Proc. Natl Acad. Sci. USA* **108,** 11959–11964 (2011).
9. Arnosti, D. N., Barolo, S., Levine, M. & Small, S. The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* **122,** 205–214 (1996).
10. Brodu, V., Mugat, B., Fichelson, P., Lepesant, J. A. & Antoniewski, C. A UAS site substitution approach to the *in vivo* dissection of promoters: interplay between the GATAb activator and the AEF-1 repressor at a *Drosophila* ecdysone response unit. *Development* **128,** 2593–2602 (2001).
11. Hens, K. *et al.* Automated protein-DNA interaction screening of *Drosophila* regulatory elements. *Nature Methods* **8,** 1065–1070 (2011).
12. Shlyueva, D. *et al.* Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. *Mol. Cell* **54,** 180–192 (2014).
13. Zabidi, M. A. *et al.* Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518,** 556–559 (2015).
14. Gerber, H. P. *et al.* Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* **263,** 808–811 (1994).
15. Galant, R. & Carroll, S. B. Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* **415,** 910–913 (2002).
16. Lam, K. C. *et al.* The NSL complex regulates housekeeping genes in *Drosophila*. *PLoS Genet.* **8,** e1002736 (2012).
17. Wright, K. J., Marr, M. T. & Tjian, R. TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc. Natl Acad. Sci. USA* **103,** 12347–12352 (2006).
18. Martens, J. A. & Winston, F. Recent advances in understanding chromatin remodeling by Swi/Snf complexes. *Curr. Opin. Genet. Dev.* **13,** 136–142 (2003).
19. Wang, Y. *et al.* IXL, a new subunit of the mammalian Mediator complex, functions as a transcriptional suppressor. *Biochem. Biophys. Res. Commun.* **325,** 1330–1338 (2004).
20. Mittler, G. *et al.* A novel docking site on Mediator is critical for activation by VP16 in mammalian cells. *EMBO J.* **22,** 6494–6504 (2003).
21. Bryant, G. O. & Ptashne, M. Independent recruitment *in vivo* by Gal4 of two complexes required for transcription. *Mol. Cell* **11,** 1301–1309 (2003).
22. Ihry, R. J. & Bashirullah, A. Genetic control of specificity to steroid-triggered responses in *Drosophila*. *Genetics* **196,** 767–780 (2014).
23. Kim, T. W. *et al.* MED16 and MED23 of Mediator are coactivators of lipopolysaccharide- and heat-shock-induced transcriptional activators. *Proc. Natl Acad. Sci. USA* **101,** 12153–12158 (2004).
24. Gan, M., Moebus, S., Eggert, H. & Saumweber, H. The Chriz-Z4 complex recruits JIL-1 to polytene chromosomes, a requirement for interband-specific phosphorylation of H3S10. *J. Biosci.* **36,** 425–438 (2011).
25. Guruharsha, K. G. *et al.* A protein complex network of *Drosophila melanogaster*. *Cell* **147,** 690–703 (2011).
26. Nibu, Y., Zhang, H. & Levine, M. Interaction of short-range repressors with *Drosophila* CtBP in the embryo. *Science* **280,** 101–104 (1998).
27. Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* **302,** 1727–1736 (2003).
28. Rhee, D. Y. *et al.* Transcription factor networks in *Drosophila melanogaster*. *Cell Rep.* **8,** 2031–2043 (2014).
29. Murali, T. *et al.* DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for *Drosophila*. *Nucleic Acids Res.* **39,** D736–D743 (2011).
30. Amelio, A. L. *et al.* A coactivator trap identifies NONO ($p54^{nrb}$) as a component of the cAMP-signaling pathway. *Proc. Natl Acad. Sci. USA* **104,** 20314–20319 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to K. Hens and B. Deplancke for sharing the TF entry clones, J. O. Yáñez-Cuna for help designing the enhancer contexts, and O. Bell, J. Brennecke, L. Cochella and S. Westermann for comments on the manuscript. We thank IMP/IMBA services, especially H. Scheuch, R. Heinen and Z. Dzupinkova, for technical support, and A. Posekany and A. Aszódi for advice on multiple-testing correction. Deep sequencing was performed at the CSF Next-Generation Sequencing Unit (http://csf.ac.at). The Stark group is supported by a European Research Council (ERC) Starting Grant (no. 242922) awarded to A.S., Boehringer Ingelheim GmbH, and the Austrian Research Promotion Agency (FFG).

Author Contributions G.S. and A.S. conceived the project. G.S. and O.F. cloned the cofactors and the GAL4–DBD fusions. G.S. and F.R. performed the luciferase assays in *Drosophila* cells and S.W. in HeLa cells. T.K. and G.S. conducted the bioinformatics analyses. G.S., T.K. and A.S. wrote the manuscript.

Additional Information All data are available at http://factors.starklab.org. The next-generation sequencing data have been deposited at the NCBI Sequence Read Archive (SRA) under the accession SRS806429. The cofactor Gateway entry clones and other plasmids are available from Addgene (http://www.addgene.org/Alexander_Stark/). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.S. (stark@starklab.org).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Cloning of N-terminal GAL4-DBD-tagged TF and cofactor library. For TFs, gateway-compatible entry clones (Invitrogen) containing the open reading frames (ORFs) lacking stop codons were obtained from ref. 11. *Drosophila Act5C*-promoter driven expression clones were created using the Gateway system. The TF ORFs were shuttled into the GAL4-DNA binding domain (DBD) containing destination vector pAGW-GAL4-DBD (cloned as described below) by mixing 100 ng of TF entry clone, 100 ng of pAGW-GAL4-DBD and 0.7 µl of LR clonase II enzyme mix (Invitrogen). The identities of all TF entry clones have been confirmed by Sanger sequencing using the primers 5′-CCCAGTCACGACGTTG-3′ and 5′-CACAGGAAACAGCTATG-3′. Note that we tested the full-length transcription factors, including their DBDs, as *trans*-activating and DNA-binding functions might not always reside in entirely separate protein domains. While this implies that the fusion proteins might bind via the TFs' DBDs in addition to the GAL4-DBD mediated recruitment, this does not influence the results of the assay: the assay itself measures transcriptional activation independently of where TF binding occurs and we expect that the TFs' DBDs have at most minor effects on binding strengths as the GAL4-DBD binds to DNA already very strongly.

For cofactors, we compiled a list of 338 cofactors based on several criteria. We included proteins containing Pfam domains typical for transcriptional cofactors (for example, HAT, HDAC, SET, Chromo, Bromo), proteins which are part of chromatin modifying or remodelling complexes or part of complexes associated with RNA polymerases (for example, SAGA, Polycomb, TFIID, Mediator), and *Drosophila* proteins which are homologues of mammalian chromatin-associated proteins (Supplementary Table 3). We amplified the cofactor ORFs from cDNA using oligonucleotides containing Gateway-compatible attB-sites (5′-GGGGACA AGTTTGTACAAAAAAGCAGGCTTC-3′ and 5′-GGGACCACTTTGTACAA GAAAGCTGGGTC-3′) for subsequent entry clone creation. The primer sequences have been chosen to be as close as possible to an annealing temperature of 60 °C which we calculated using the formula $T_{\text{ann}} = \frac{64.9 + 41 \times (\text{cG} + \text{cC} - 16.4)}{(\text{cA} + \text{cT} + \text{cG} + \text{cC})}$ with cA, cT, cG and cC being the number of adenines, thymines, guanines and cytosines, respectively. The full list of resulting primer sequences (lacking the attB sequences) is listed in Supplementary Table 3; for 18 of the cofactors no primer sequences are available because we obtained these entry clones from ref. 11 categorized as TFs but manually re-categorized them as cofactors based on their annotation in FlyBase[31] or their protein domain content[32]. For cDNA generation, RNA was isolated from S2 cells and reverse transcribed as described in ref. 33. For PCR amplification, KOD and KOD XL DNA (Merck Millipore) and KAPA HiFi (KAPA) polymerases were used according to manufacturer's specifications. We created Gateway entry clones by mixing 1 µl of PCR reaction, 100 ng of pDONR221, and 1 µl of BP clonase II enzyme mix (Invitrogen). The identities and correctness of all entry clones have been ensured using Sanger and next-generation sequencing (see below) and we deposited them at Addgene (http://www.addgene.org/Alexander_Stark/). The cofactor ORFs were then shuttled to the *Drosophila Act5C*-promoter driven destination vector pAGW-GAL4-DBD as described for TFs.

Verification of cofactor clones by Sanger and next-generation sequencing. The insert flanks of all obtained cofactor entry clones have been Sanger-sequenced and automatically checked to cover the TSS and TTS of one of the isoforms annotated by FlyBase. All entry clones passing additional manual visual inspection using BLAT and the UCSC genome browser have been subjected to further verification by next-generation sequencing as follows. A pool of 100–300 entry clones corresponding to a total of 5 µg DNA solved in 50 µl TE buffer was sonicated (duty cycle, 20%; intensity, 5; 200 cycles per burst; time, 90 s) to 200–400 bp using a S220 Focused-ultrasonicator (Covaris) as described in ref. 11. The fragmented plasmid pool was then prepared for deep sequencing using the Illumina DNA Sample Prep kit and sequenced using a HiSeq2000 (Illumina) producing 50-nt reads. The resulting reads have been assembled and analysed using PrInSes-C[34]. All insert sequences not starting with ATG, containing a stop codon or a frameshift were immediately rejected. All sequences with less than five mutations leading to non-synonymous amino acid changes were immediately accepted. The remaining sequences were translated, aligned against the respective protein sequence, and manually decided. The next-generation sequencing reads have been deposited at the NCBI Sequence Read Archive (SRA) under the accession SRS806429; the PrInSes-C-generated full-length transcript sequences are available at http://factors.starklab.org and in Supplementary Data 1, and the cofactor Gateway entry clones from Addgene (http://www.addgene.org/Alexander_Stark/).

Cloning of destination vector pAGW-GAL4-DBD. We cloned a destination vector to conveniently create vectors expressing N-terminally V5- and GAL4-DBD-tagged TFs and cofactors under the control of the *Drosophila Act5C* promoter using the Gateway cloning system. pAGW-GAL4-DBD was cloned by amplifying the GAL4-DBD from pBPGUw[35] using one oligonucleotide containing the V5-tag (peptide sequence MGKPIPNPLLGLDST) 5′-TCTGATATCATGGGGAAGCC AATCCCTAATCCCCTTCTGGGACTCGACTCTACCGGCGGCTCTATGAA GCTACTGTCTTCTATCGAACA-3′ and the oligonucleotide 5′-TATACCGGT GGCCGCCGCCCGACGATACAGTCAACTGTCTTTGAC-3′. Amplification was performed using KOD Polymerase (Merck Millipore) according to the manufacturer's instructions. The resulting PCR product was digested using EcoRV and AgeI and ligated into pAGW (*Drosophila* Gateway Vector Collection), which was digested using the same enzymes, thereby replacing eGFP with V5-GAL4-DBD.

Cloning of luciferase reporter vectors. We created Gateway-compatible (Invitrogen) destination vectors to conveniently clone reporter vectors for different regulatory contexts based on firefly luciferase transcribed from a housekeeping core promoter (hkCP; promoter of ribosomal gene *RpS12*[13]) or a developmental core promoter (dCP; *Drosophila* synthetic core promoter (DSCP) derived from *Eve*[35]).

We created the destination vector attR_dCP_luc by digesting pGL4.26 (Promega) with FseI and BglII and ligating a fragment containing DSCP and luc+, thereby replacing the minimal promoter and luc2 with DSCP-luc+. We digested the resulting vector with KpnI and BglII and ligated a fragment containing the attR Gateway cassette, yielding attR_dCP_luc. We created two hkCP-driven destination vectors containing a Gateway cassette either upstream (attR_hkCP_luc) or downstream (hkCP_luc_attR) of the luciferase reporter gene by using the plasmid pGL3 (Promega) as a basis and replacing the SV40 promoter with the promoter of *RpS12* as described in ref. 13. The resulting vector was digested using either KpnI and BglII (to create attR_hkCP_luc) or AfeI (to create hkCP_luc_attR); in both cases, we amplified a Gateway attR cassette using oligonucleotides containing the respective restriction sites, and digested and ligated it into the digested plasmid.

All enhancers, motif mutant contexts and other motif or backbone mutant variants were either PCR amplified with primers containing attB Gateway sites or ordered as synthesized fragments (IDT), shuttled into entry clones using TOPO or BP Clonase II (both Invitrogen), and shuttled into the luciferase destination vectors using the LR clonase II enzyme mix (Invitrogen) by mixing 1 µl of PCR product or synthesized DNA solved in TE buffer, 100 ng of destination vector and 0.7 µl of LR clonase II enzyme mix (Invitrogen).

We used a modified version of pRL-TK (Promega) to normalize the firefly signal for transfection efficiency and cell number. Ubi-RL has been created by cloning a region upstream of the gene *Ubi-p63E* (chr3L: 3901760-3902637) upstream of the *Renilla* luciferase gene in reverse orientation using NheI and BglII.

***Drosophila* cell culture.** S2 cells, derived from embryos[36], were obtained from Life Technologies and grown in Schneider's *Drosophila* Medium (Life Technologies 21720-024) supplemented with 10% FBS (Sigma F7524) and 1% penicillin/streptomycin (Life Technologies 15140-122) grown in T75 flasks (ThermoScientific 156499) at 27 °C and passaged every 2–4 days. BG3 neuroblast-like cells, derived from larvae[37], were obtained from the *Drosophila* Genomics Resource Center (DGRC) and grown in Schneider's *Drosophila* Medium supplemented with 10% FBS, 1% penicillin/streptomycin, and 10 µg ml^{-1} Insulin (Sigma-Aldrich I1882) in T75 flasks at 27 °C and passaged every 3–4 days. Kc167 cells, derived from embryos[38], were obtained from DGRC and grown in M3/BPYE Medium containing 5% FBS and 1% penicillin/streptomycin in T75 flasks at 27 °C and passaged every 2–3 days. Ovarian somatic cells (OSCs), derived from adult ovaries[39], were obtained from the laboratory of J. Brennecke and grown in Shields and Sang M3 Insect Medium (Sigma-Aldrich S8398) supplemented with 10% FBS, 1% insulin, 1% glutathione, 1% fly extract, and 1% penicillin/streptomycin in T75 flasks at 27 °C and passaged every 2–3 days. All cell lines used are regularly checked for mycoplasma contamination.

Transfections of *Drosophila* cell lines. S2 cell transfections were performed using jetPEI (peqlab 13-101-40N). Four hours before transfection, 30,000 cells (30 µl of a 10^6 cells per ml suspension) were seeded in clear polystyrene 384-well plates (ThermoScientific 164688). For each transfection, we used 30 ng firefly luciferase reporter plasmid, 3 ng *Renilla* luciferase expressing plasmid Ubi-RL, and 3 ng GAL4-DBD-TF/cofactor or GAL4-DBD-GFP fusion protein expressing plasmid. Beforehand, we assayed the effects of using different amounts of GAL4-DBD fusion protein expressing plasmid and chose 3 ng (Extended Data Fig. 9). The DNA solution containing 36 ng DNA in 5 µl TE buffer was filled up to 15 µl using sterile 150 mM NaCl (polyplus) and prepared in 96-well plates. Transfection reagent (15 µl total: 13.95 µl 150 mM NaCl, 1.05 µl jetPEI) was added to each well of the 96-well plates and mixed rigorously. After 30 min incubation at 25 °C, cells were transfected in quadruplicates by transferring each transfection mix four times (6 µl each) to four adjacent wells of a 384-well plate containing the seeded cells. Luciferase assays were performed after 48 h of growth at 27 °C. Handling the transfection mixes and all subsequent pipetting steps have been performed using a Bravo Automated Liquid Handling Platform (Agilent). Kc167, BG3, and

OSC cell transfections were performed using jetPEI in the same way as described above for S2 cells with the exception of transfection reagent composition: 15 µl total containing 14.1 µl 150 mM NaCl and 0.9 µl jetPEI.

HeLa cell culture and transfections. Human HeLa cells (gift from the laboratory of J. M. Peters) were grown in DMEM medium (Gibco 52100-047) supplemented with 10% heat-inactivated FBS, 1% penicillin/streptomycin and 2 mM L-glutamine (Sigma G7513) in T75 flasks at 37 °C in an atmosphere of 95% air and 5% carbon dioxide. All cell lines used are regularly checked for mycoplasma contamination. We performed HeLa cell transfections using a self-prepared 1 mg ml^{-1} PEI (25,000 MW, Polysciences 23966) stock solution in PBS (pH adjusted to pH 4.5 and sterile filtered). On the day before transfection we seeded 30 µl of a suspension containing 4,000 HeLa cells in medium (DMEM, 10% FBS, penicillin/streptomycin) into each well of a 384-well plate. Three microlitres of a PEI/DMEM mix (0.24 µl PEI filled to a total of 4.5 µl using DMEM without FBS and penicillin/streptomycin and incubated at room temperature for 5 min) were added to 3 µl of a DNA/DMEM mix (44.5 ng firefly luciferase reporter vector, 4.45 ng TF expression vector (created using pAGW-CMV_ GAL4-DBD, see below) and 4.45 ng pRL-CMV vector for transfection normalization (Promega #E2261) in DMEM without FBS and penicillin/streptomycin. The resulting DNA/PEI mix in DMEM was incubated at room temperature for 30 min and subsequently added to the seeded cells. We performed cell lysis and luciferase assays using the Promega dual-luciferase reporter assay system (Promega E1910) according to the manual.

We created the Gateway destination vector pAGW-CMV_GAL4-DBD by replacing the *Drosophila Act5C* promoter in pAGW-GAL4-DBD with a region containing the CMV enhancer and the T7 promoter amplified from pRL-CMV using the primers 5′-CGACAGATCTTCAATATTGGCCATTAGCCATAT-3′ and 5′-GGTGGCTAGCCTATAGTGAGTCGTATTA-3′.

Luciferase assays. Dual-luciferase assays were performed using self-prepared substrate solutions (D-Luciferin and Coelenterazine have been obtained from GoldBio LUCK-250 and pjk-Gmbh 102111) and lysis buffer as described in ref. 40. For cell lysis, the supernatant was removed and 30 µl of lysis buffer added and incubated gently shaking for 30 min. Ten microlitres of the cell lysates were transferred to black 384-well plates for luminescence assays (Nunc MaxiSorp, Sigma-Aldrich P6491-1CS). All pipetting steps have been performed using a Bravo Automated Liquid Handling Platform (Agilent). Luminescence was measured after adding 20 µl of each substrate, for firefly and *Renilla* luciferase respectively, using a Biotek Synergy H1 plate reader coupled to a plate stacker.

Luciferase data analysis and plots. We normalized all firefly luciferase signals to the signal of *Renilla* luciferase to control for transfection efficiency and cell number (the relative luciferase signal). We then further normalized all relative luciferase signals for TF- and cofactor-GAL4-DBD transfections to relative luciferase signals obtained for GAL4-DBD-GFP transfections (fold-change over GFP). We assessed statistical significance by two-sided unpaired *t*-tests on the two sets of quadruplicate relative luciferase signals (GAL4-DBD-TF/COF versus GAL4-DBD-GFP). Throughout the paper, 'activation' was defined as a fold-change ≥ 1.5 ($P < 0.05$), and 'repression' was defined as a fold-change $\leq 1/1.5$ ($P < 0.05$), both compared to the signal for GAL4-DBD-GFP. We corrected the *P* values for multiple testing using the Benjamini and Hochberg method as implemented in R (p.adjust with method 'BH' or its alias 'fdr'). All statistical calculations and graphical displays, if not stated otherwise, have been performed using version 2.15.3 of the R software suite[41].

TF cluster feature enrichment analysis. Enrichment analyses have been performed for each of the 15 clusters and for 6 types of features. To first obtain a coarse functional characterization of the clusters, we assessed the enrichments and depletions of TFs which are able to activate or repress a developmental (dCP) or housekeeping (hkCP) core promoter on their own (≥ 1.5-fold activation or repression ($P < 0.05$), both compared to the signal for GAL4-DBD-GFP when tested on a context comprised of UAS sites upstream of a developmental core promoter (4×UAS-dCP) or a housekeeping core promoter hkCP (4×UAS upstream hkCP)). Homopolymeric amino acid repeat motifs have been *de novo* discovered using MEME[42] (version 4.8.1, *q*-value threshold of 1×10^{-5}) in TFs that activated or repressed on their own outside enhancer contexts (tested in the 4×UAS dCP context; ≥ 1.5-fold; $P < 0.05$). Pfam domain[32] signature matches in the *Drosophila* proteome have been generated using hmmer[43] (version 3.0b3, *e*-value threshold of 0.01). Eukaryotic Linear Motifs[44] (ELM; version 08/2014) were matched to the amino acid sequences of the tested TF protein isoforms, after masking the TFs' Pfam. Additionally, Gene Ontology[45] (GO) annotations, and gene expression patterns in the *Drosophila* embryo as annotated by ref. 46 (IMAGO) have been subjected to enrichment and depletion analyses.

To control for multiple testing, we empirically determined false-discovery rates (FDRs) for the different hypergeometric *P* values. For this, we repeated the feature enrichment analyses 1,000 times, each after randomly shuffling the TF-to-cluster assignments, and recorded the best (that is, most significant) *P* values. We then adjusted the original *P* values such that only 10% of the 1,000 random controls reached the *P* values of the original data (FDR < 10%). Following this protocol, we separately adjusted the FDR cut-off for each cluster (15) and feature type (ELM, MEME, Pfam, GO, IMAGO).

Validation with the TFs' endogenous motifs. To assess if tethering via the GAL4-DBD reflects the different TFs' regulatory functions when bound to their endogenous motifs, we selected two sets of TFs, three TFs that preferentially activated the CGCG- versus the GATA-context (Fig. 1e) and four TFs that preferentially activated the hormone-receptor contexts; Fig. 2c). We replaced each UAS site in the enhancer mutant contexts S2-1 CGCG, S2-1 GATA, and Nhe2 EcR[3,12] (which also corresponds to an endogenous TF motif in the wild-type enhancers, for example, the EcR motif for the hormone contexts) with a sequence corresponding to the consensus motif of the respective TF as reported in refs 47, 48. (Dfd: CTTAATGA, Hey: CAGCCGACACGTGCCCC, Ets21C: ATTTCCGGT, Ato: AACAGGTGG, Ets96B: ACCGGAAGTAC, Gl: ATTTCAAGAATA, HLH4C: AAAAACACCTGCGCC). The enhancer rescue constructs were synthesized by IDT, shuttled into the luciferase reporter vector attR_dCP_luc using the Gateway system and tested in luciferase assays in S2 cells exactly as described above.

TF–cofactor association assays. To assess potential functional associations of assigned TFs and cofactors, we followed the strategy from ref. 30, recruiting TFs via GAL4-DBD and providing untagged cofactors. For this, we chose contexts in which the different TFs (Clk of cluster 8, Bsh of cluster 10, and CG17186 of cluster 14) were active (4×UAS-dCP for Clk and 4×UAS-upstream-hkCP for Bsh and CG17186). We prepared DNA mixes to be transfected containing 29 ng firefly luciferase reporter plasmid, 3 ng *Renilla* luciferase expressing plasmid Ubi-RL, 1 ng (Bsh and CG17186) or 0.5 ng (Clk) of GAL4-DBD–TF fusion protein expressing plasmid and an increasing series of untagged cofactor expressing plasmid (0 ng, 0.003 ng, 0.006 ng, 0.012 ng, 0.023 ng, 0.047 ng, 0.094 ng, 0.188 ng, 0.375 ng, 0.75 ng, 1.5 ng, 3 ng). We kept the total amount of transfected plasmid DNA constant at 36 ng for all experiments using a GFP-expressing plasmid. To clone the expression plasmids for the untagged cofactors and GFP, we used the Gateway-compatible vector pAW (*Drosophila* Gateway Vector Collection). The remaining experimental procedure and analysis was performed as described above.

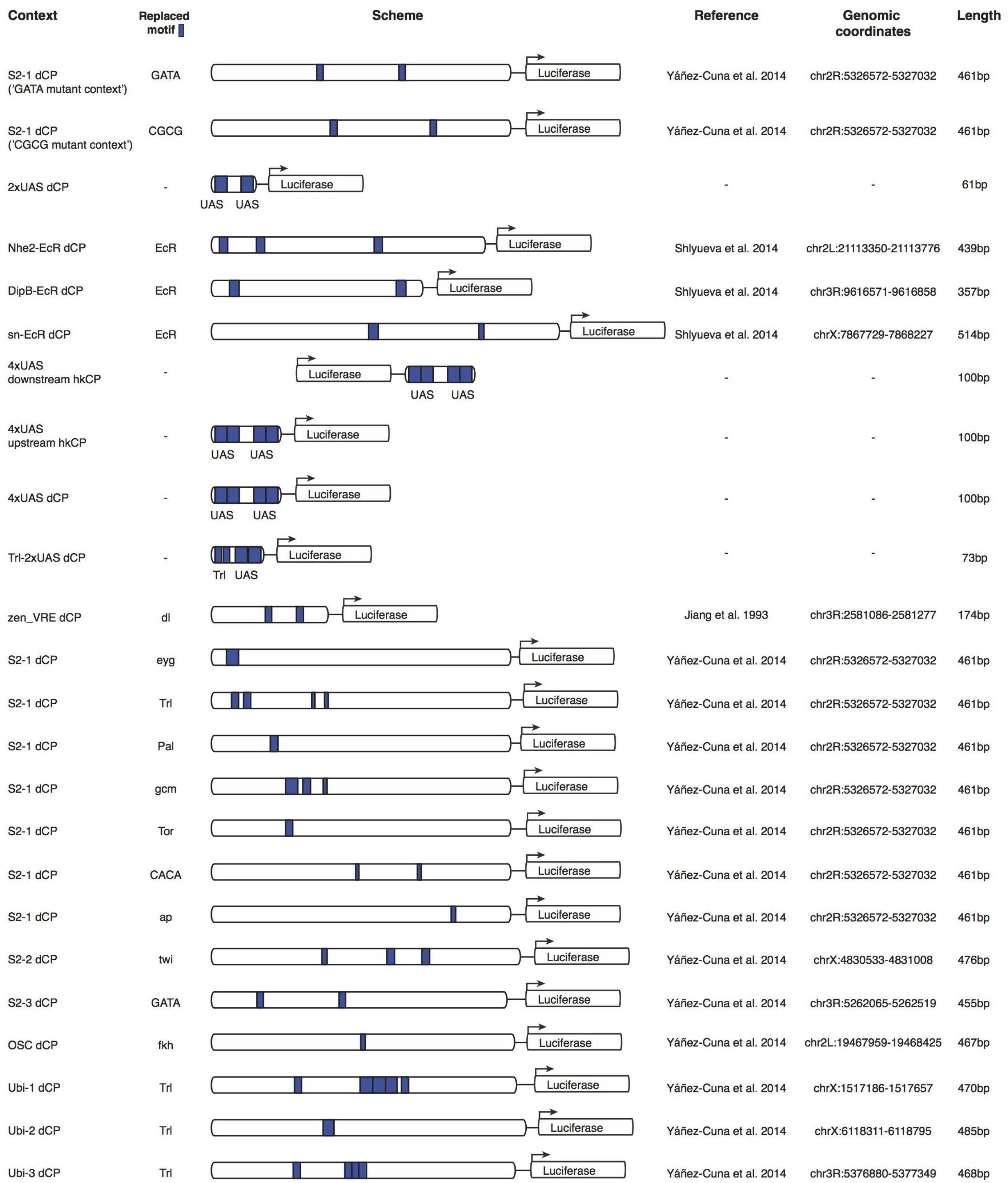
Transcription factor clustering, visualization, and assignment of cofactors to transcription factors. We clustered the 474 TFs based on the $\log_2$-transformed fold-change values (TF over GFP) from all 24 contexts. First, we standardized all contexts and constructed a *k*-nearest-neighbour graph ($k = 15$). We used the Euclidean distance as distance measure as it reflects both the variation of the enhancer activity profile across contexts and the effect sizes within each context; that is, it is able to discriminate between strong and weak activators and repressors even if they vary similarly across the 24 contexts. Next, we took a symmetrized ($A + A^T$) adjacency matrix of this graph and solved multiclass spectral clustering as described in ref. 49 and implemented in the Python package scikit-learn[50]. In order to decide about the number of clusters and to assess the clustering validity, we analysed the clustering stability upon bootstrapping the data set[51]. In order to visualize the data, we mapped the data onto a plane by a specialized nonlinear dimensionality reduction technique (t-SNE)[52]. The algorithm provides the visualization by mapping data points close in the original space to nearby locations in the plane, preserving the local structure. We extended the *k*-nearest-neighbour graph to include cofactors by comparing the $\log_2$-transformed fold-change values (cofactor over GFP) of cofactors and TFs ($k = 5$, Euclidean distance). The locations of the cofactors in the visualization were obtained from spring layout.

TF candidate recovery of enhancer mutants. We know that UAS sites in the enhancer mutant contexts most probably replace binding sites that are functional[3] but we do not know which TFs bind them *in vivo*. In order to check whether we recover these positive controls in the enhancer mutants, we took all the TFs expressed in S2 cells (RPKM > 1) (ref. 53) for which motifs are known[3]. We scanned the wild-type enhancer sequences (S2-1-wt, S2-2-wt, S2-3-wt, Ubi-1-wt, Ubi-2-wt, Ubi-3-wt) for motif matches with $P < 9.76 \times 10^{-4}$ (1/4,096) using an in-house motif-detection program. For each mutant context, we considered only those TFs for which any of its motif matches had at least 5 mutated base pairs. In the resulting set of TFs (Extended Data Table 1) there is at least one TF per each of the enhancer mutant contexts that activated the respective context when recruited via the GAL4-DBD (≥ 1.5-fold activation compared to GFP; $P < 0.05$).

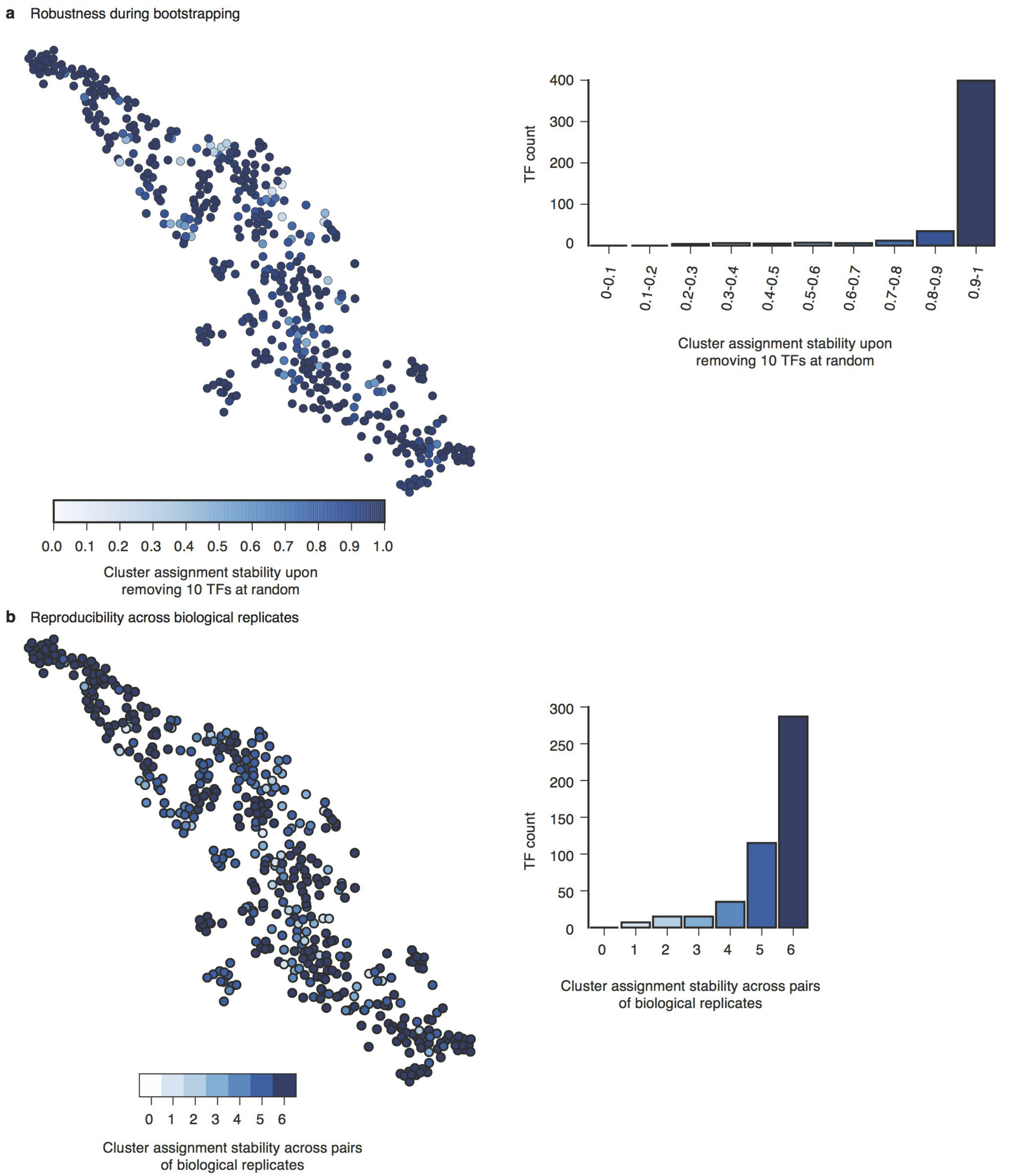
Cell type analysis—distances intra- versus inter-cluster. We tested a subset of the original 472 TFs in four different cell types (S2, Kc167, BG3 and OSC). This subset consists of 171 TFs covering all the 15 clusters by 9–17 TFs, including all the TFs mentioned in the main text. In each cell type, we computed Euclidean distances after standardizing the $\log_2$-transformed fold-change values in each context. Then we compared the distances of intra-cluster TF–TF pairs (both TFs belong to the same cluster) to inter-cluster TF–TF pairs (each of the TFs belongs to a different

cluster). In order to test whether the medians of these two groups of distances are significantly different, we determined empirical *P* values as follows. We randomly shuffled TF-to-cluster assignments 10^6 times and each time computed the medians of the distances for both groups. We mark the *P* values $P < 1 \times 10^{-6}$ as we never obtained a difference between the medians of intra- and inter-cluster distances as large as for the actual data for any of the cell types.

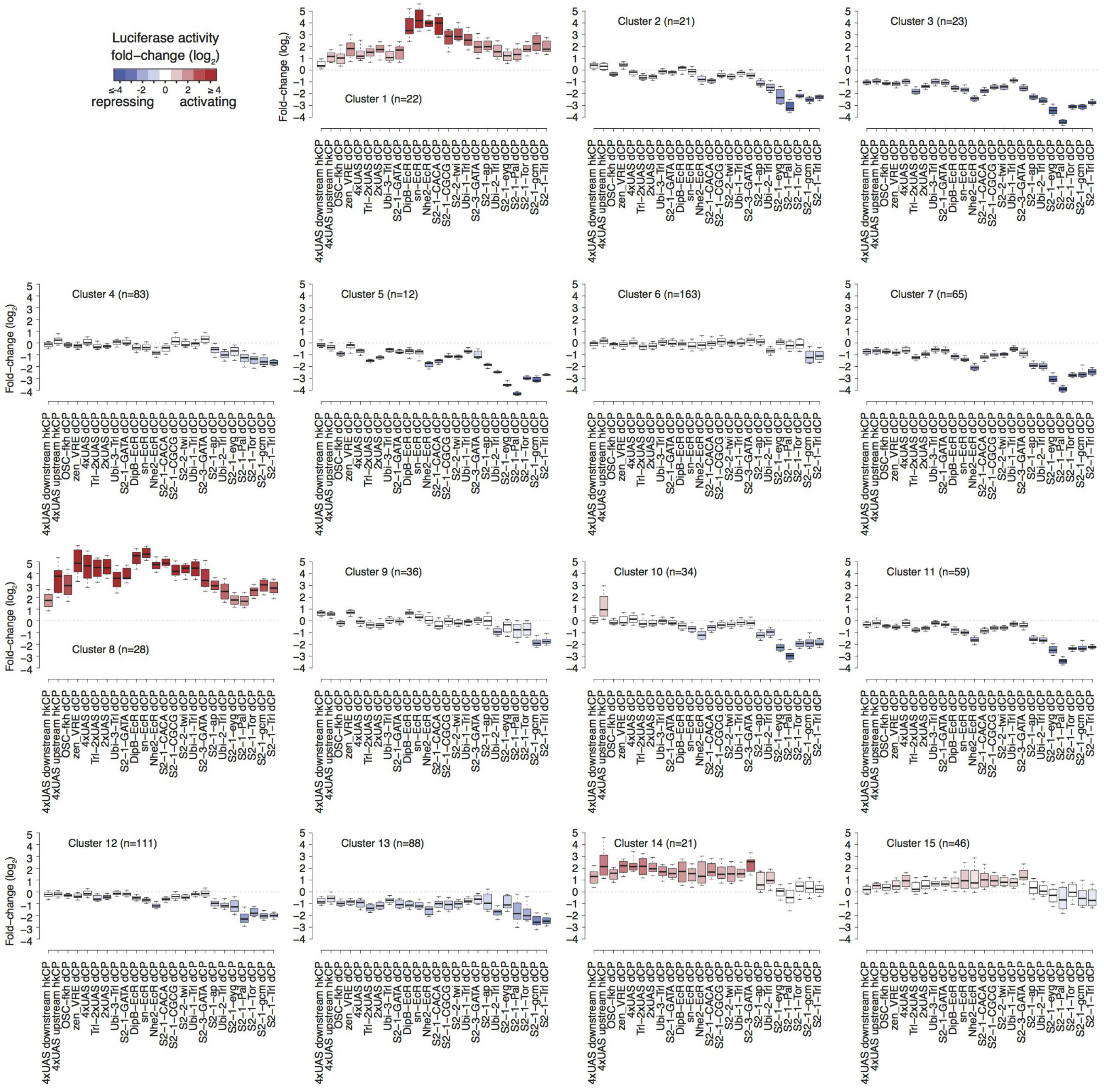
31. St Pierre, S. E., Ponting, L., Stefancsik, R. & McQuilton, P. FlyBase Consortium. FlyBase 102–advanced approaches to interrogating FlyBase. *Nucleic Acids Res.* **42,** D780–D788 (2014).
32. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42,** D222–D230 (2014).
33. Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339,** 1074–1077 (2013).
34. Massouras, A., Decouttere, F., Hens, K. & Deplancke, B. WebPrInSeS: automated full-length clone sequence identification and verification using high-throughput sequencing data. *Nucleic Acids Res.* **38,** W378–W384 (2010).
35. Pfeiffer, B. D. *et al.* Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proc. Natl Acad. Sci. USA* **105,** 9715–9720 (2008).
36. Schneider, I. The culture of cells from insects and ticks. I. Cultivation of dipteran cells *in vitro*. *Curr. Top. Microbiol. Immunol.* **55,** 1–12 (1971).
37. Ui, K. *et al.* Newly established cell lines from *Drosophila* larval CNS express neural specific characteristics. *In Vitro Cell. Dev. Biol. Anim*. **30,** 209–216 (1994).
38. Echalier, G. & Ohanessian, A. Isolation, in tissue culture, of *Drosophila melangaster* cell lines. *C. R. Acad. Sci. Hebd. Seances Acad. Sci. D* **268,** 1771–1773 (1969) [transl.].
39. Saito, K. *et al.* A regulatory circuit for *piwi* by the large Maf gene *traffic jam* in *Drosophila*. *Nature* **461,** 1296–1299 (2009).
40. Hampf, M. & Gossen, M. A protocol for combined *Photinus* and *Renilla* luciferase quantification compatible with protein assays. *Anal. Biochem.* **356,** 94–99 (2006).
41. R Development Core Team. *R: A language and environment for statistical computing.* (*R Foundation for Statistical Computing*, 2011).
42. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37,** W202–W208 (2009).
43. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14,** 755–763 (1998).
44. Dinkel, H. *et al.* ELM–the database of eukaryotic linear motifs. *Nucleic Acids Res.* **40,** D242–D251 (2012).
45. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25,** 25–29 (2000).
46. Tomancak, P. *et al.* Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **3,** (2002).
47. Zhu, L. J. *et al.* FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res.* **39,** D111–D117 (2011).
48. Stark, A. *et al.* Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450,** 219–232 (2007).
49. Yu, S. X. & Shi, J. Multiclass spectral clustering. In *ICCV*, 313–319 (IEEE, 2003).
50. Pedregosa, F. *et al.* Scikit-learn: machine learning in Python. *J. Machine Learn. Res.* **12** 2825–2830 (2011).
51. von Luxburg, U. Clustering stability: an overview. *Foundations and Trends in Machine Learning* **2,** 235–274 (2010).
52. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Machine Learn. Res.* **9,** 2579–2605 (2008).
53. The modENCODE Consortium *et al*. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330,** 1787–1797 (2010).
54. Pfeiffer, B. D. *et al.* Refinement of tools for targeted gene expression in *Drosophila*. *Genetics* **186,** 735–755 (2010).



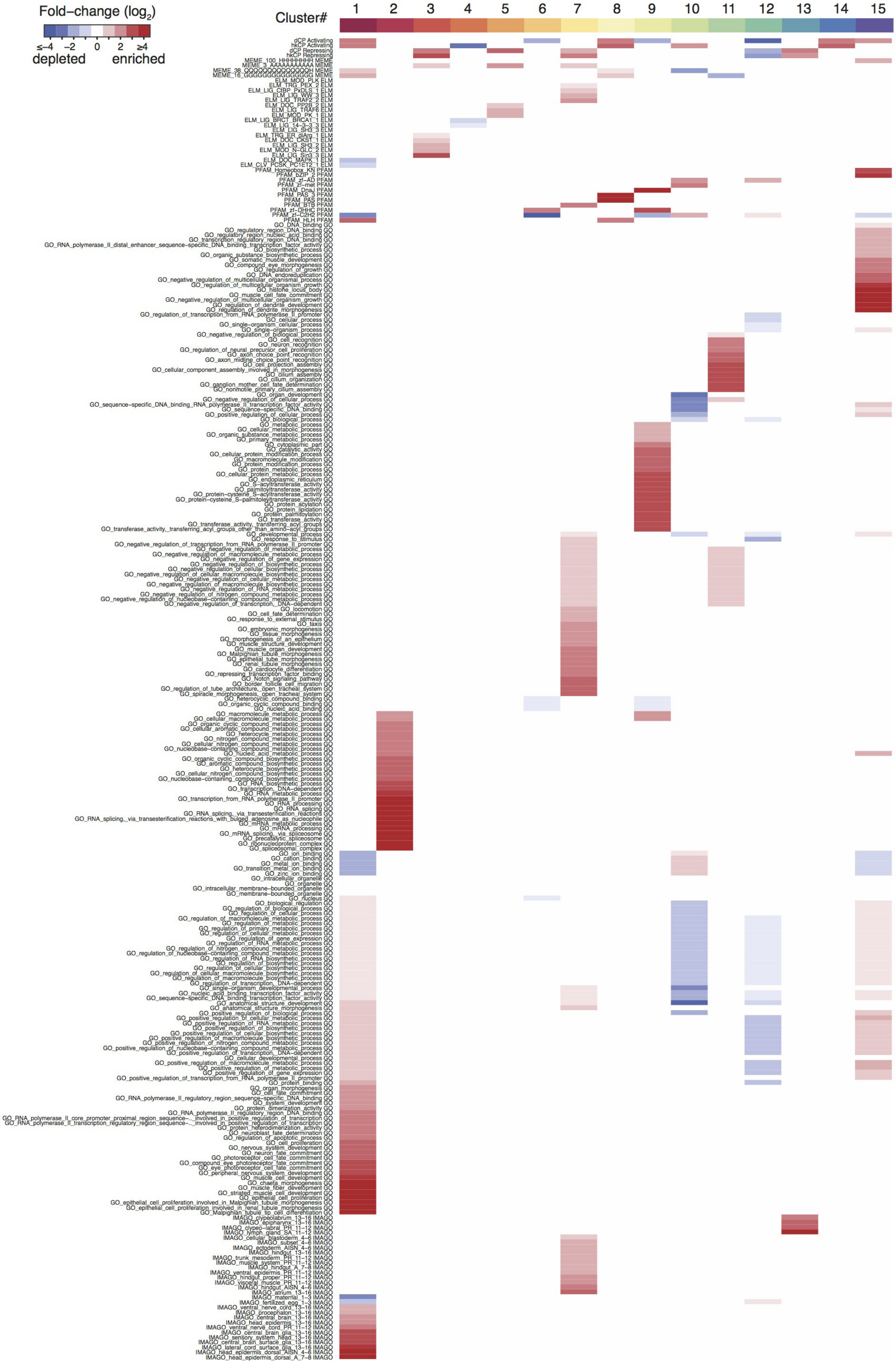
Extended Data Figure 1 | 24 regulatory contexts. Tested contexts included 19 motif-mutant enhancer contexts which were designed by replacing 43 occurrences of 15 different motif types in 11 previously characterized enhancers with broad ('Ubi-1' to 'Ubi-3') or cell type-specific ('S2-1' to 'S2-3': S2 cell-specific; 'OSC': ovarian somatic cell (OSC)-specific) activities (all from ref. 3) or hormone-inducible enhancers (from ref. 12.). We also designed five synthetic contexts consisting of UAS sites with or without Trl sites and a developmental core promoter (dCP; *Drosophila* synthetic core promoter (DSCP) derived from the transcription factor gene *Eve*[54]) or with a housekeeping core promoter (hkCP; derived from the ribosomal gene *RpS12*[13]). Shown are schemes of the luciferase reporter constructs used for the targeted recruitment of GAL4-DBD–TF/cofactor fusion proteins to UAS sites (the luciferase gene is not drawn to scale). Motif names denote the motifs (as named by refs 3, 12) that have been replaced by UAS sites (blue boxes) to create the enhancer context. Note that TF-to-motif assignments are not unique and typically several TFs can bind each of the motifs (see Extended Data Table 1).



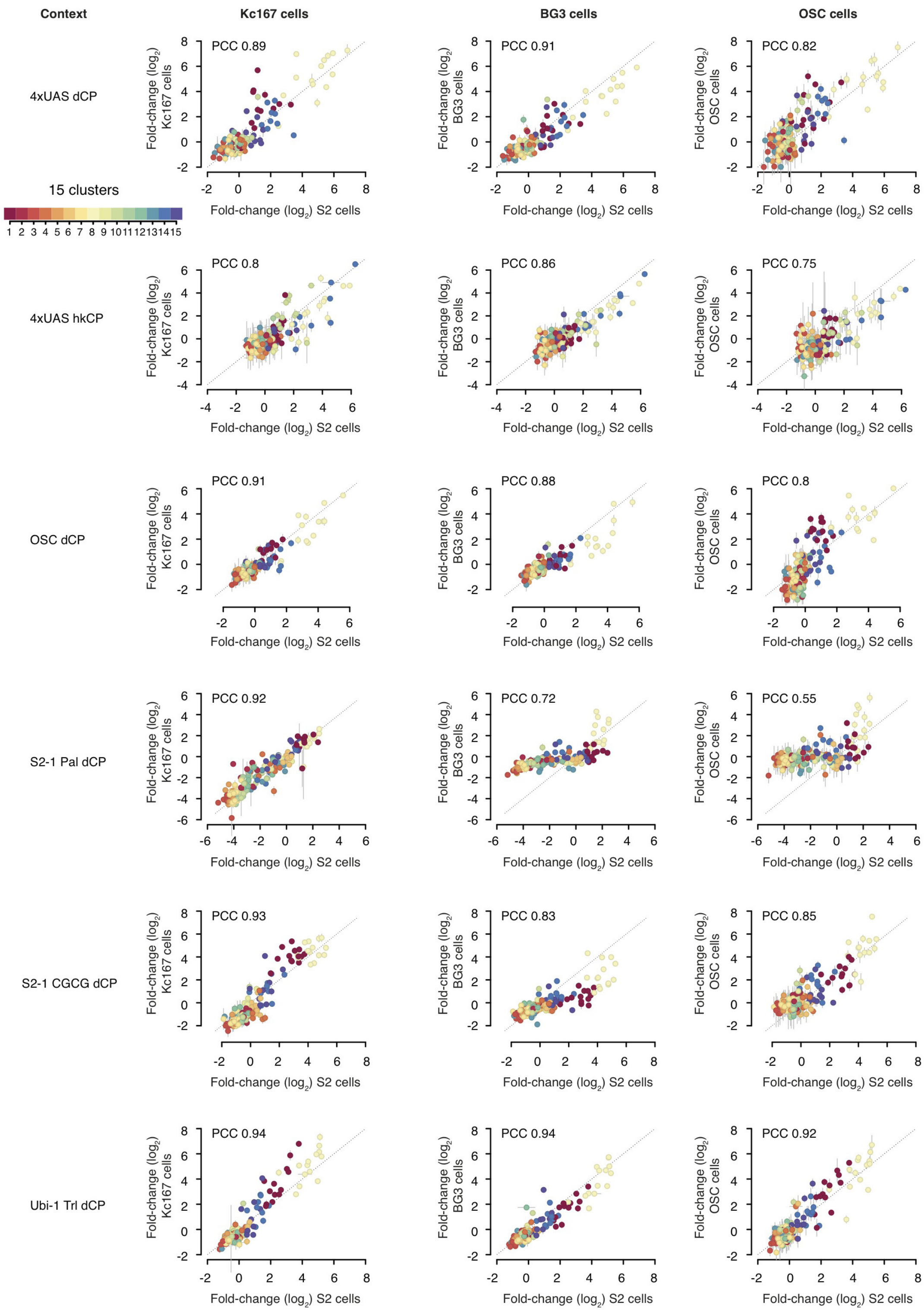
Extended Data Figure 2 | TF clustering is robust and reproducible. **a**, Cluster assignment is robust during bootstrapping (474 rounds of removing 10 randomly selected TFs). The cluster label stability denotes the fraction (out of 474 trials) a given TF was assigned to the same cluster as in the original clustering (node layout shown is identical to Fig. 2a). The vast majority of TFs were assigned to the same group in ≥90% of the cases (histogram). **b**, Cluster assignment for individual TFs is reproducible for biological replicates. We repeated the clustering six times, each time using only two out of the four biological replicates (six corresponds to all possibilities to choose two out of the four replicates). The cluster stability denotes the number of times (out of six) a given TF is assigned to the same cluster as in the original clustering. The majority of TFs were assigned to the same cluster, independent of which pair of biological replicates was used to generate the clustering (histogram).



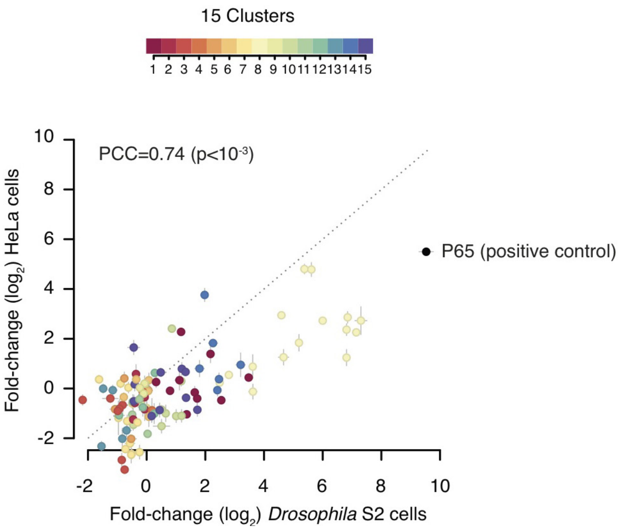
Extended Data Figure 3 | Cluster activity profiles. Normalized luciferase values for all TFs assigned to each of the 15 clusters across all 24 contexts. Shown are median and quartiles as boxes, and the tenth and ninetieth percentiles as whiskers for each of the 24 contexts. Boxes are coloured according to the median activity in each context (see colour legend).



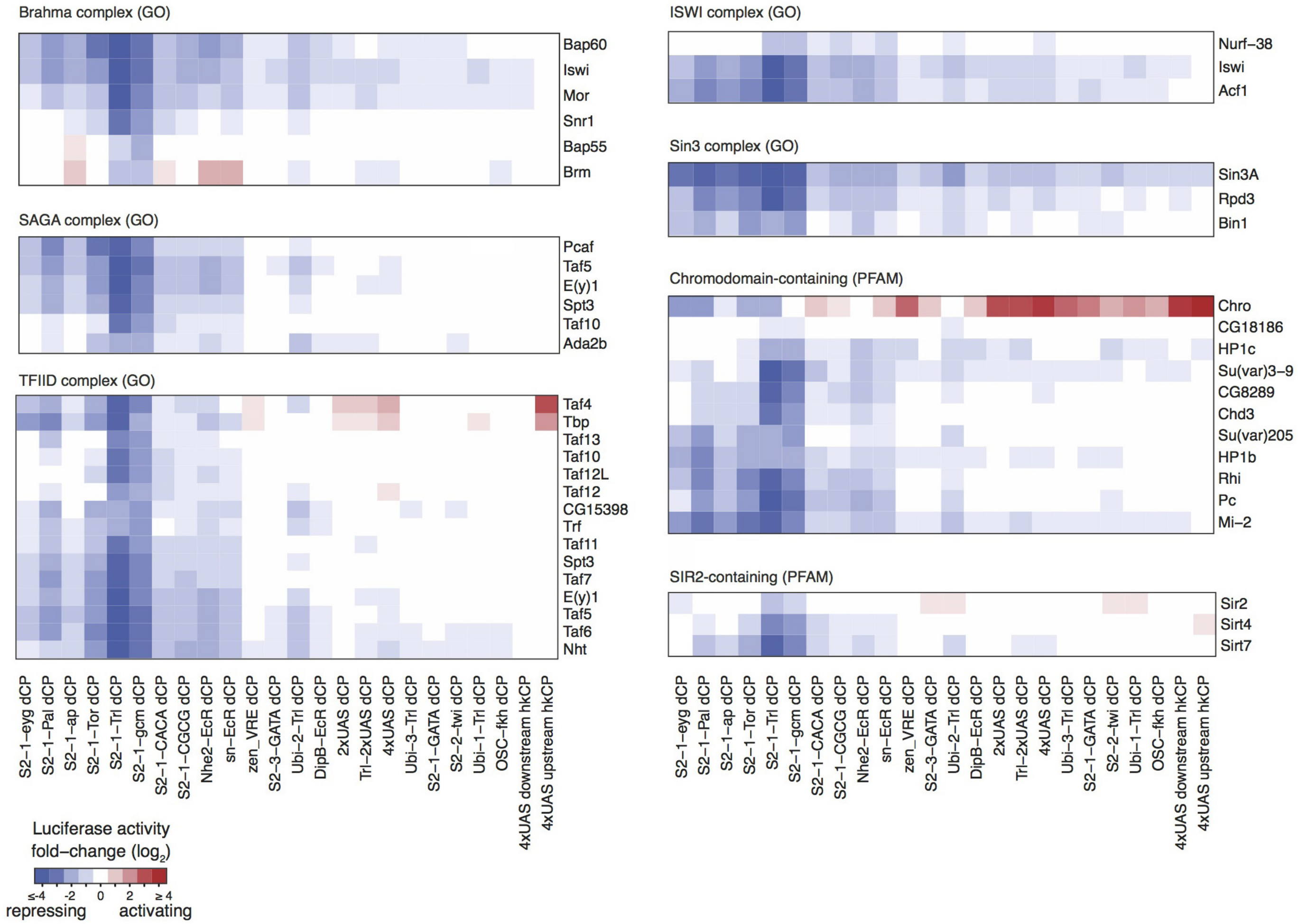
Extended Data Figure 4 | Complete feature enrichment analysis of 15 groups of TFs. Features analysed include eukaryotic linear motifs[44] (ELM), homopolymeric amino acid repeat motifs discovered using MEME[42], protein domains as annotated by Pfam[32], Gene Ontology[45] (GO), and gene expression patterns as annotated by IMAGO[46]. Red and blue shadings denote enrichments and depletions (log$_2$-transformed) with an empirical FDR of at most 10% per feature type and cluster (others are white).



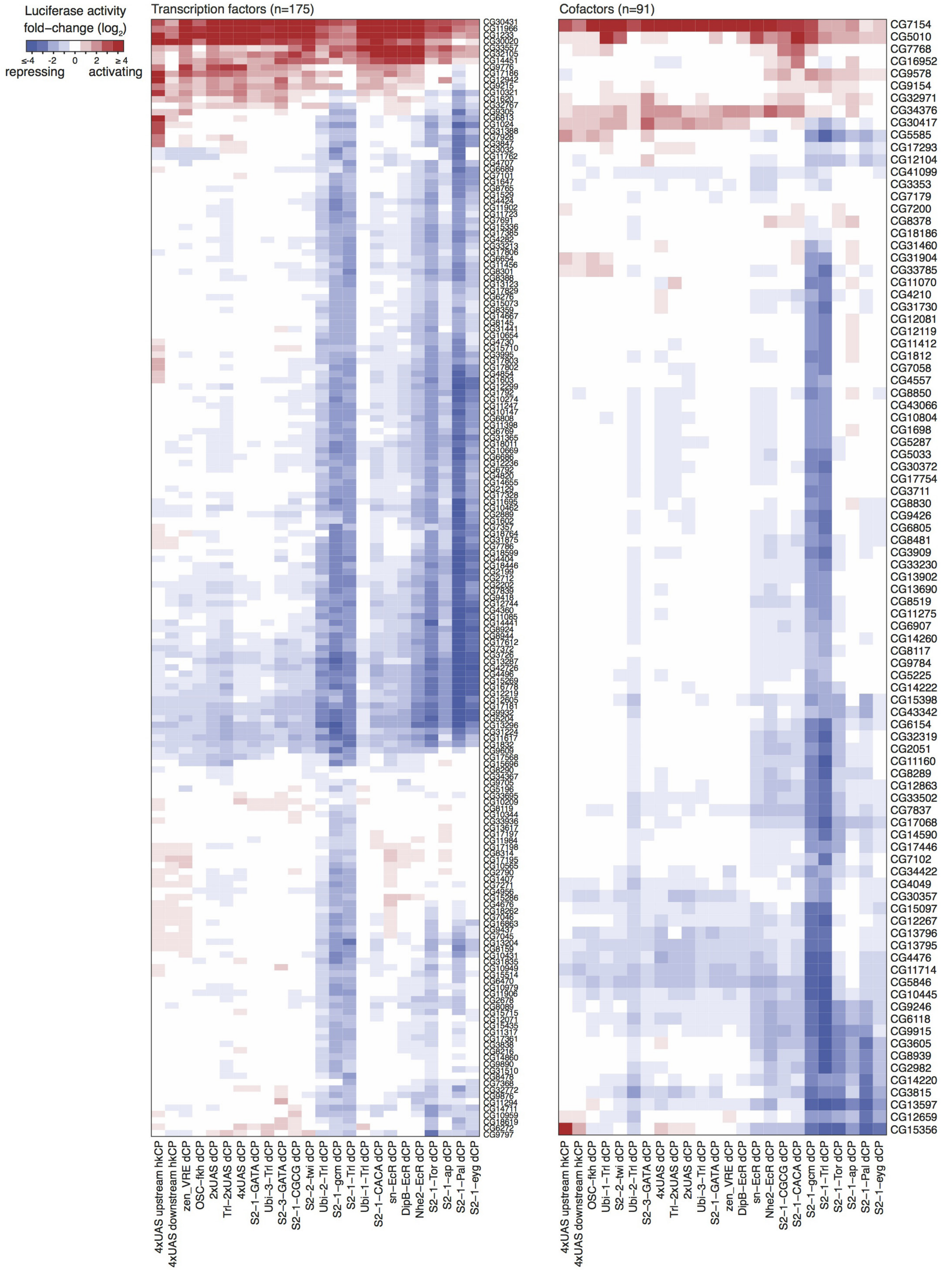
Extended Data Figure 5 | TFs behave consistently across *Drosophila* cell types. We tested 171 of the 474 TFs (36.1%) in 6 of the 24 contexts in Kc167, BG3 and OSC cell lines, which are derived from embryos, larvae and adult, respectively. Shown are normalized luciferase values and the Pearson correlation coefficients (PCC; $P < 1 \times 10^{-3}$ for all comparisons). We tested synthetic contexts containing an array of UAS sites upstream of a developmental and a housekeeping core promoter[13] (4×UAS dCP and hkCP), three contexts derived from cell-type-specific enhancers[3] (OSC dCP, S2-1 Pal dCP, S2-1 CGCG dCP), and one context derived from a broadly active enhancer[3] (Ubi-1 Trl dCP). The latter showed the highest similarities (PCCs of 0.94, 0.94 and 0.92 for Kc167, BG3 and OSC cells, respectively) while the lowest PCCs for the non-embryonic BG3 and OSC cells (0.72 for BG3 and 0.55 for OSC) were obtained for S2-1 Pal dCP, derived from an enhancer active only in the embryonic S2 and Kc167 cells, presumably because the corresponding wild-type enhancer sequence is inactive in larval and adult cells[3,12] such that combinatorial effects between the tethered TF and other enhancer-bound TFs may be less effective or lack entirely. Enhancer complementation presumes (and the results throughout this study confirm this presumption) that the regulatory functions of the tethered TFs are revealed (or altered) by other enhancer-bound factors; that is, factors that are bound to the enhancer in S2 cells (in which the corresponding enhancer is active[3]) but not in the other cell types (in which the enhancer is not active). This emphasizes the value of enhancer complementation for the study of regulatory activities and the importance of contexts derived from active enhancers. Error bars denote standard deviation ($n = 4$, biological replicates).



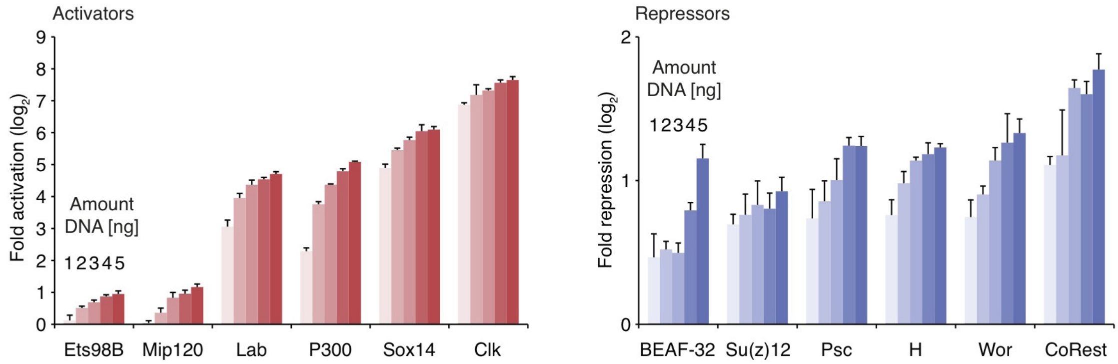
Extended Data Figure 6 | *Drosophila* TFs and cofactors retain their activating functions in human HeLa cells. We expressed GAL4-DBD fusion proteins for 107 of the 812 *Drosophila* factors (90 TFs and 17 cofactors) under the control of a constitutively active CMV promoter in human HeLa cells (see Methods). Shown are normalized luciferase values for the tested proteins recruited to the synthetic 4×UAS-dCP context. The values are remarkably similar quantitatively, with an overall Pearson correlation coefficient (PCC) of 0.74 ($P < 1 \times 10^{-3}$). The activation domain of the human TF P65 was used as a positive control. Error bars denote standard deviation ($n = 4$, biological replicates).



Extended Data Figure 7 | Regulatory activities of selected cofactor complexes or protein domain families. Heat maps of normalized luciferase values for sets of proteins annotated as being part of the same complex by Gene Ontology[45] (GO) or containing a chromodomain or SIR2 domain as annotated by Pfam[32].



Extended Data Figure 8 | Regulatory activities of uncharacterized TFs and cofactors. Heat maps of normalized luciferase values for all 'CG genes' among the tested TFs and cofactors, which activate or repress in at least one context (≥1.5-fold compared to GFP; $P < 0.05$ FDR-corrected for 24×474 and 24×338 tests for TFs and cofactors, respectively).



Extended Data Figure 9 | Consistent effects of varying the amounts of plasmid DNA for TF and cofactor expression. The effects of using 1 ng, 2 ng, 3 ng, 4 ng and 5 ng of GAL4-DBD–TF/cofactor fusion protein expressing plasmids on luciferase assays in S2 cells suggest that reporter activity is robust to variation in TF levels. Shown are normalized luciferase values of GAL4-DBD N-terminally fused to six activating and six repressing TFs and cofactors of different strengths targeted to the synthetic 4×UAS-dCP context. The amount of plasmid expressing the GAL4-DBD–TF/cofactor fusion proteins was 3 ng for all factors throughout this study. Error bars denote standard deviation ($n = 4$, biological replicates).

Extended Data Table 1 | TF recovery analysis for S2 cell enhancer contexts

CONTEXT	TF	RPKM_S2-CELLS	FOLD ACTIVATION
S2-1-CACA dCP	Cf2	15.16	2.24
S2-1-CGCG dCP	vfl	13.54	27.04
S2-1-GATA dCP	Cf2	15.16	2.42
S2-1-GATA dCP	sd	109.42	1.53
S2-1-GATA dCP	sqz	12.39	1.68
S2-2-twi dCP	Cf2	15.16	1.81
S2-2-twi dCP	gl	24.29	5.55
S2-2-twi dCP	twi	92.36	5.82
S2-3-GATA dCP	ap	644.03	1.80
S2-3-GATA dCP	Atf6	17.04	6.60
S2-3-GATA dCP	lab	1.44	7.69
Ubi-1-Trl dCP	Atf6	17.04	2.75
Ubi-1-Trl dCP	Hsf	38.62	6.39
Ubi-1-Trl dCP	Sox14	17.55	29.68
Ubi-2-Trl dCP	Hsf	38.62	3.14
Ubi-3-Trl dCP	Hsf	38.62	5.00

TFs with known motifs that match the sequences we mutated to UAS sites in each of the different enhancer contexts are expressed in S2 cells (RPKM > 1 (ref. 53)) and significantly activate the respective context when recruited via the GAL4-DBD (≥1.5-fold activation compared to GFP; $P < 0.05$).

CORRIGENDUM

doi:10.1038/nature16075

Corrigendum: Regulatory analysis of the *C. elegans* genome with spatiotemporal resolution

Carlos L. Araya, Trupti Kawli, Anshul Kundaje, Lixia Jiang, Beijing Wu, Dionne Vafeados, Robert Terrell, Peter Weissdepp, Louis Gevirtzman, Daniel Mace, Wei Niu, Alan P. Boyle, Dan Xie, Lijia Ma, John I. Murray, Valerie Reinke, Robert H. Waterston & Michael Snyder

Nature **512,** 400–405 (2014); doi:10.1038/nature13497

In this Article, when processing *C. elegans* ChIP-seq libraries, the gene label ZK337.2 (KLU-1, a C2H2 Zn-finger protein) was mis-transcribed to ZK377.2 (SAX-3), a neuronal fate regulator. To clarify, ZK337.2 (KLU-1) is not an established neuronal fate regulator, but joins FKH-10 and C34F6.9 as an unstudied gene grouped with previously established neuronal regulators (SEM-4, MAB-5, CES-1 and ZAG-1). This error affects Figs 1g and 2, Extended Data Figs 3–7 and 10, and Supplementary Tables 1 and 4 of the original Article. In addition, KLU-1, not SAX-3, changes from neuronal targets in L2 larvae to carbohydrate/lipid metabolism targets in L4 larvae.

CORRIGENDUM

doi:10.1038/nature16136

Corrigendum: Mutant IDH inhibits HNF-4α to block hepatocyte differentiation and promote biliary cancer

Supriya K. Saha, Christine A. Parachoniak, Krishna S. Ghanta, Julien Fitamant, Kenneth N. Ross, Mortada S. Najem, Sushma Gurumurthy, Esra A. Akbay, Daniela Sia, Helena Cornella, Oriana Miltiadous, Chad Walesky, Vikram Deshpande, Andrew X. Zhu, Aram F. Hezel, Katharine E. Yen, Kimberly S. Straley, Jeremy Travins, Janeta Popovici-Muller, Camelia Gliser, Cristina R. Ferrone, Udayan Apte, Josep M. Llovet, Kwok-Kin Wong, Sridhar Ramaswamy & Nabeel Bardeesy

Nature **513,** 110–114 (2014); doi:10.1038/nature13441
corrigendum *Nature* **519,** 118 (2015); doi:10.1038/nature14149

In Extended Data Fig. 1b of this Letter, the photomicrographic images of the hepatoblast cells grown under normal conditions were mismatched. The figure shows control images indicating that cells expressing mutant IDH1 (R132C and R132H) or mutant IDH2 (R140Q and R172K) have similar morphology to those expressing wild-type (WT) IDH1 or IDH2 or empty vector (EV). The errors in the figure were: in the top row, the EV and R132C panels were swapped and in the bottom row, the panel for R172K was replaced by a duplicate image of the IDH2 WT panel. The Supplementary Information for this Corrigendum contains the corrected Extended Data Fig. 1b, and the original source files from which the corrected figure was assembled. Our conclusions are unaffected.

Supplementary Information is available in the online version of the Corrigendum.

CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature16157

Corrigendum: The formation and fate of internal waves in the South China Sea

Matthew H. Alford, Thomas Peacock, Jennifer A. MacKinnon, Jonathan D. Nash, Maarten C. Buijsman, Luca R. Centurioni, Shenn-Yu Chao, Ming-Huei Chang, David M. Farmer, Oliver B. Fringer, Ke-Hsien Fu, Patrick C. Gallacher, Hans C. Graber, Karl R. Helfrich, Steven M. Jachec, Christopher R. Jackson, Jody M. Klymak, Dong S. Ko, Sen Jan, T. M. Shaun Johnston, Sonya Legg, I-Huan Lee, Ren-Chieh Lien, Matthieu J. Mercier, James N. Moum, Ruth Musgrave, Jae-Hun Park, Andrew I. Pickering, Robert Pinkel, Luc Rainville, Steven R. Ramp, Daniel L. Rudnick, Sutanu Sarkar, Alberto Scotti, Harper L. Simmons, Louis C. St Laurent, Subhas K. Venayagamoorthy, Yu-Huai Wang, Joe Wang, Yiing J. Yang, Theresa Paluszkiewicz & Tswen-Yung (David) Tang

Nature **521,** 65–69 (2015); doi: 10.1038/nature14399

In this Letter, the surname of author Luca Centurioni was incorrectly spelt Centuroni; this has been corrected in the online versions of the paper.

TOOLBOX

ANNOTATING THE SCHOLARLY WEB

Scientific publishers are forging links with an organization that wants scientists to scribble comments over online research papers.

ILLUSTRATION BY THE PROJECT TWINS

BY JEFFREY M. PERKEL

Would researchers scrawl notes, critiques and comments across online research papers if software made the annotation easy for them? Dan Whaley, founder of the non-profit organization Hypothes.is, certainly thinks so.

Whaley's start-up company has built an open-source software platform for web annotations that allows users to highlight text or to comment on any web page or PDF file. And on 1 December, Hypothes.is announced partnerships with more than 40 publishers, technology firms and scholarly websites, including Wiley, CrossRef, PLOS, Project Jupyter, HighWire and arXiv.

Whaley hopes that the partnerships will encourage researchers to start annotating the world's online scholarship. Scientists could scribble comments on research papers and share them publicly or privately, and educators could use annotation to build interactive classroom lessons, he says. If the idea takes off, some enthusiasts suggest that the ability to annotate research papers online might even change the way that papers are written, peer reviewed and published.

Hypothes.is, which was founded in 2011 in San Francisco, California, and is supported by philanthropic grants, has a bold mission: "To enable conversations over the world's knowledge." But the concept it implements, online annotation, is as old as the web itself. The idea of permitting readers of web pages to annotate them dates back to 1993; an early version of the Mosaic web browser had this functionality. Yet the feature was ultimately discarded. A few websites today have inserted code that allows annotations to be made on their pages by default, including the blog platform Medium, the scholarly reference-management system F1000 Workspace and the news site Quartz. However, annotations are visible only to users on those sites. Other ▸

annotation services, such as A.nnotate or Google Docs, require users to upload documents to cloud-computing servers to make shared annotations and comments on them.

Hypothes.is is not the only service that wants to make it easy for users to leave annotations across the entire web. A competing offering is a web annotation service from Genius, a start-up firm that began as a site for annotating rap lyrics. In April, it launched services such as browser plugins to help users to annotate any web page. But unlike Hypothes.is, the Genius code is not open-source, its service doesn't work on PDFs, and it is not working with the scholarly community. On the scholarly side, the reference-management tool ReadCube makes it possible for users to annotate PDFs of papers viewed on a ReadCube web reader — but that software is proprietary. (ReadCube is owned by Digital Science, a firm operated by the Holtzbrinck Publishing Group, which also has a share in *Nature*'s publisher.)

By contrast, the open-source nature of the Hypothes.is platform means that anyone could use it to create their own annotation reader or writer — just as anyone can create their own web browser using standards-based technology. The company is also a member of a working group within the World Wide Web Consortium, the standards body for the web, which is developing a universal standard for annotations and how they are communicated. The hope is that web pages that allow annotations would all adopt the same underlying code and protocols (as they do with hyperlinks, for example), making the function easier to use and interact with. The working group has released a draft version of its standard, which is expected to be finalized by the end of 2016.

HOW IT WORKS

For now, Hypothes.is users have several options for creating and viewing annotations. These include bookmarklets (a simple program within a browser bookmark), browser plugins or adding 'via.hypothes.is/' to the start of any URL.

When a Hypothes.is user opens a page — a scholarly article, for instance — the web browser shows any annotations to which the user has access. These appear as highlighted words and comments on top of the text, like an overlaid transparency. Users can then add their own comments, similar to a student marking up a textbook. These are public by default but can be made private, and, following an update added on 3 November, annotations can be shared with private groups. That should enable the tool to be used for journal clubs, classroom exercises and even peer review.

If a page has been altered since an annotation was made, the software uses 'fuzzy' logic to map annotations to their approximate original location. The system can also map annotations from HTML to PDF and back again (for instance, if a user annotates the web version of an article and subsequently views a PDF of the same document).

Annotations are stored on a dedicated Hypothes.is server, which Whaley says looks set to log around 250,000 comments from some 10,000 users in 2015. For instance, after Hurricane Patricia in October, climate scientists left comments and highlighted text on a widely shared mashable.com article (see go.nature.com/rcsesf). But publishers that wish to host annotations for their own content, or companies that want to annotate corporate documents behind a firewall, could run their own server using the same software platform, Whaley adds.

PUBLISHER PARTNERSHIPS

A Hypothes.is user can already annotate any web page — including research papers and pay-to-view articles to which they have access. But the formal partnership announced this week sees some publishers working harder to encourage annotation, including tackling content that annotation systems stumble over, such as page frames and embedded page readers.

The digital library JSTOR, for example, is developing a custom Hypothes.is tool for its educational project with the Poetry Foundation, a literary organization and publisher in Chicago, Illinois. Alex Humphreys, who is director of JSTOR Labs in New York City, says that teachers will be able to use the tool to annotate poems with their classes. An instructor selects the poem to be annotated, sets up a dedicated page with a copy of it, and restricts access to their class only. Students can then create personal notes or share them with the group; an extra annotation layer finds the scholarly resources in JSTOR that quote each line of poetry and provides links out to those resources. The tool is slated to launch in mid-December, Humphreys says.

> ***"You can think of this as a fabric that allows those comments to move freely in time and [across] versions."***

The scientific publisher eLife in Cambridge, UK, has been testing the feasibility of using Hypothes.is to replace its peer-review commenting system, says Ian Mulvany, who heads technology at the firm. The publisher plans to incorporate the annotation platform in a site redesign instead of its current commenting system, Disqus. At a minimum, says Mulvany, Hypothes.is provides a mechanism for more-targeted commentary — the equivalent of moving comments up from the bottom of a web page into the main body of the article itself.

Another partner, the arXiv preprint service run by Cornell University Library in Ithaca, New York, has been working on making annotations flow across multiple article versions, says information scientist Simeon Warner, who leads technology development for arXiv. To jump-start interest in the annotation program, arXiv has been converting mentions of its articles in external blog posts (called trackbacks) into annotations that are visible on an article's abstract page when using Hypothes.is.

NOT JUST GRAFFITI

Hypothes.is plans improvements to its platform that include a way to validate the identities of commenters, by incorporating researchers' unique ORCID digital profiles. That could go a long way towards improving adoption of the system among scholars, by facilitating expert commentary on published works and filtering out unwanted marginalia, says Paul Ginsparg, the founder of arXiv and a physicist at Cornell University. "If people start looking at articles and they see the equivalent of graffiti, then people will turn off the comments and the experiment will fail," he says.

If it takes off, online annotation could represent a fundamental shift in the way scholarly communication is done, adds Cameron Neylon, part of the research team at the Centre for Culture and Technology at Curtin University in Perth, Australia, who formerly worked at PLOS.

At the moment, Neylon explains, the scholarly publishing process involves ferrying a document from place to place. Researchers prepare manuscripts, share them with colleagues, fold in comments and submit them to journals. Journal editors send copies to peer reviewers, returning their comments to the author, who goes back and forth with the editor to finalize the text. After publication, readers weigh in with commentary of their own.

With an open-source annotation platform, Neylon says, the document is the centre of attention. Different contributors act on the content simply by changing who has access to it and its comments, with the document becoming richer over time. "You can think of this as a fabric that allows those comments to move freely both in time and [across] versions in a way that we've never been able to do before," he says.

But as Ginsparg points out, it is not clear that researchers — who have proved reluctant in repeated trials to comment on published articles — will take to annotation, even if they can share their comments privately. "There's no incentive structure for people to comment extensively, because it can take time to write a thoughtful comment, and one currently doesn't get credit for it," he says. "But it's an experiment that needs to be done." ■

Jeffrey M. Perkel *is a writer based in Pocatello, Idaho.*

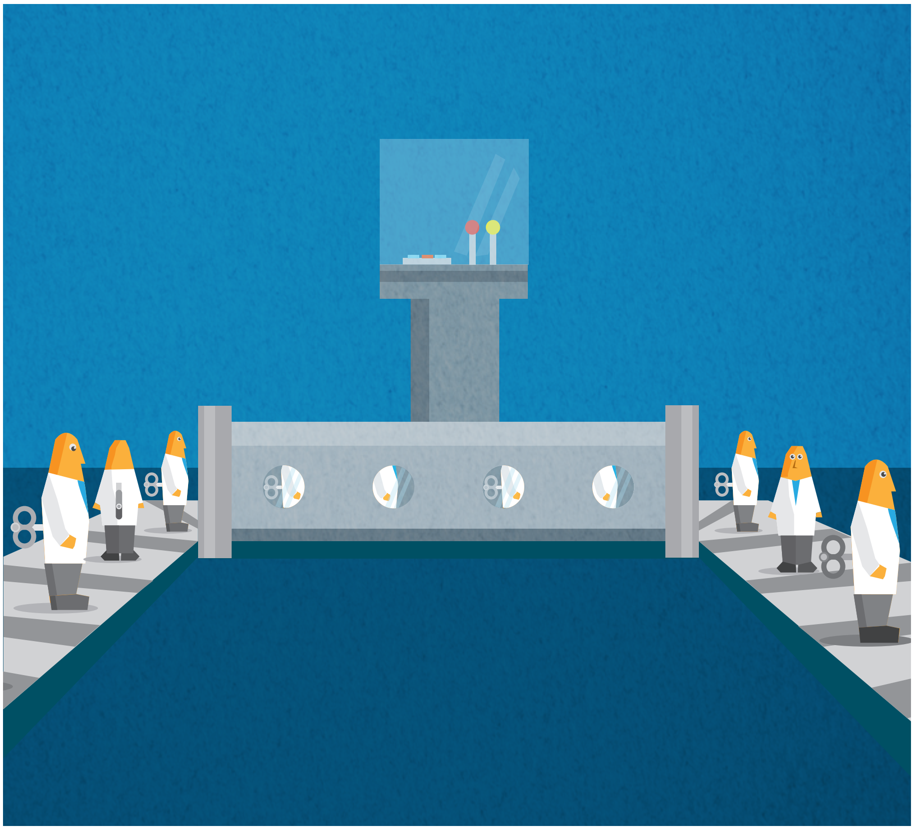
CORRECTION

The table in the Toolbox article 'Eight ways to clean a digital library' (*Nature* **527,** 123–124; 2015) wrongly stated that ReadCube runs on only desktop and web platforms. In fact, it also runs on mobile platforms.

POSTDOCS UNITE New rights for California trainees **p.156**

TRANSPARENCY Overcoming the data-sharing challenge **go.nature.com/rwwpai**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**



CLAIRE WELSH/NATURE

COLUMN

Fellowships are the future

Postdocs need a level of autonomy to get the best out of their position, say **Viviane Callier** and **Jessica Polka**.

Much scientific research could not function without postdocs. They do the research outlined in a grant — moving the work of the principal investigator (PI) forward, producing papers and helping to win grants. Yet too many postdocs end up doing work that does not benefit their scientific and intellectual development. They are shut out of developing ownership of a research programme, a step that is crucial for launching the next stage of their career.

Many have argued that postdocs should be classified as employees so that they can receive compensation and benefits for their many hours of labour. Yet a postdoc is not simply an employee, providing a service in exchange for money. Postdoctoral training is an important window during which a researcher can pick up new skills and ideas that will help her or him to establish an independent laboratory or move into a permanent post.

Often, those goals align well with those of the PI, and the two reinforce each other. But when funding is tight and push comes to shove, postdocs may end up advancing their PI's career without developing the necessary skills to progress their own. This misalignment has only widened as the academic job market has become more competitive and research funding scarcer.

To address US postdocs' need for career development, the White House Office of Management and Budget last year issued a statement asserting that postdocs are both employees and trainees and confirming that they should have protected time to pursue career-development activities. The statement is a step in the right direction, but it lacks teeth. Means for enforcing the policy are unclear, and many scientists hold the unrealistic expectation that a postdoc will spend 100% of his or her time working on the PI's grant, despite the need to prepare for the next career step.

We argue that funding agencies should support postdocs through fellowships. Currently, only about 16% of US postdocs are supported by training grants or fellowships; the rest are paid through funding that is awarded to their PI. We advocate viewing the postdoctoral stint as a transitional period during which the postdoc develops independence. Such a shift would be made possible if postdocs were funded directly, rather than through a PI's grant as are most technicians and staff scientists. Financial security helps to foster intellectual independence — and fellowships are better positioned to provide that. Fellowships also provide more security because they can be transferred with the postdoc should she or he change institutions, and they guarantee the postdoc a specific number of years of funding. When a postdoc is funded through a PI's grant, those grants may have to be cobbled together over several years.

ADVANTAGES AND CHALLENGES

In our experience, fellowships have several advantages over more-conventional funding routes. Training can be built into the experience more easily, for example, and applying for them provides the opportunity to outline a research programme and to seek out a team of mentors, both at their own institution and at others. That gives postdocs an opportunity to try their hand at a key component of the work required of an assistant professor.

A mandate for postdocs to be supported through fellowships would help funding agencies to track and regulate closely the ▸

number of trainees in the system. The application and review process would also help to ensure that only researchers with the ability and desire to become independent investigators receive the awards. If few awards were available, they would probably become extremely competitive and could significantly reduce the number of postdocs in the pipeline, at least in the United States.

However, to serve as a practical alternative, these fellowships must provide postdocs with the same workplace benefits that grant-supported postdocs enjoy. That is not uniformly the case. For example, in the United States, some fellowship recipients are left to shop for health insurance on their own. When one of us (V.C.) moved from a university-sponsored salaried position to a fellowship, she lost her health insurance through the university and had to find another provider. Although the 2010 passage of the US Affordable Care Act has made health care more affordable for individuals, such a demand poses an unnecessary burden on postdocs. Instead, why not simply give all postdocs access to the same health benefits that graduate students or conventional university employees receive?

There are other challenges: boosting the number of fellowships would also increase the burden on the grant-review system. And, because research grants would not cover postdoc salaries, the arrangement could leave PIs in a precarious position — with the funds to pay for research supplies and equipment, but fewer incentives to offer postdocs to join their lab. PIs would also no longer be able to hire postdocs to cover the roles of technicians and staff scientists.

> ***"Having postdocs strike out away from the beaten path will bring fresh ideas and approaches to the table."***

If postdocs receive greater independence, PIs will lose some control, so they may have to find other resources to conduct their research. But this could be good for science: having postdocs strike out away from the beaten path will bring fresh ideas and approaches to the table. For both of us, getting a fellowship enabled us to cut a path that was separate from the dominant research area in each of our mentors' labs. The experience of trying to define a new scientific direction has been most useful for us, even as our paths diverge. ■ SEE NEWS FEATURE P. 22

Viviane Callier *is a freelance science writer and contractor at the US National Cancer Institute in Bethesda, Maryland.* **Jessica Polka** *is a postdoctoral researcher at Harvard Medical School in Boston, Massachusetts.*

EMPLOYMENT TERMS

California postdocs win new rights

Career development a priority in university contract.

BY HELEN SHEN

JENS MAGNUSSON/GETTY

Following recommendations from august US groups including the National Academies of Sciences, Engineering, and Medicine and the National Institutes of Health, some universities are initiating policies to give postdoctoral researchers more time and latitude for career exploration.

A contract between the University of California (UC) and its postdoc labour union gives postdocs across the ten-campus system the right to pursue career-development activities on paid time. "That's a pretty big milestone," says Belinda Huang, executive director of the US National Postdoctoral Association in Washington DC. "UC is being very explicit here, where other universities have not necessarily been explicit."

The agreement, which follows months of negotiation and covers the university's roughly 6,500 postdocs, took effect on 1 November, and will last until 30 September 2016. It also includes new employment protections for international postdocs and establishes a committee to consider financial assistance for child-care expenses.

The university's move echoes national-level support for such activities. In 2014, the White House Office of Management and Budget stated that it recognizes the 'dual role' of postdocs as both employees and trainees and that it expects them to be "actively engaged in their training and career development" while conducting research supported by government grants.

TIME TROUBLES

But it remains to be seen whether these measures will spur greater day-to-day participation in the courses, mentoring programmes and other career-exploration resources that already exist at many US universities, including the UC. For UC postdocs who had previously skipped such programmes because of pressure to focus on a conventional academic career, the new contract could provide an encouraging nudge. For others, however, even contract-enshrined 'permission' may not be enough. "The truth is, I don't have time to do all of these activities," says neuroscientist Wan-Yu Hsu, a postdoc at UC San Francisco. Hsu says that the drive to produce results and to publish will probably keep her from pursuing career-development activities during the workday.

The contract also addresses immigration issues. Specifically, it offers added protections for international postdocs who get fired. The university already has a grievance process for postdocs to contest terminations. But under US immigration law, many international postdocs must return to their home countries immediately on termination, leaving them to argue their cases from afar.

The UC agreement states that if the grievance process cannot be resolved at an earlier point, and the postdoc has had to leave the country, the university will help to sponsor a travel visa for the postdoc to return to the United States to participate in their final arbitration hearing. If the postdoc is successful in their case, the university will reimburse the travel costs. If he or she is not successful, the union will foot the bill.

Roughly two-thirds of UC postdocs come from outside the United States, but terminations are rare, and those that are not resolved before arbitration are even rarer, says Anke Schennink, president of the postdoc union. Nevertheless, the provision adds a measure of security, she says.

One of the outstanding issues in the contract is whether and how the university will help with child-care costs. A handful of institutions — including Stanford University in California, Cornell University in Ithaca, New York, and Princeton University in New Jersey — offer subsidies or discounts for child care. But the expenses remain a serious issue for many others across the country, and contribute to women leaving the scientific workforce, says Huang.

The postdoc union had pushed for the creation of a financial-assistance programme to help postdocs with child-care expenses — the university already offers the benefit to graduate students — but the two parties could not agree on terms. Instead, the university and union have agreed to form a committee to discuss the issue in the coming year. "We're hoping to make more progress in the next round," says Schennink. ■

BEYOND 550 ASTRONOMICAL UNITS

The joys of planet-spotting.

BY MIKE BROTHERTON

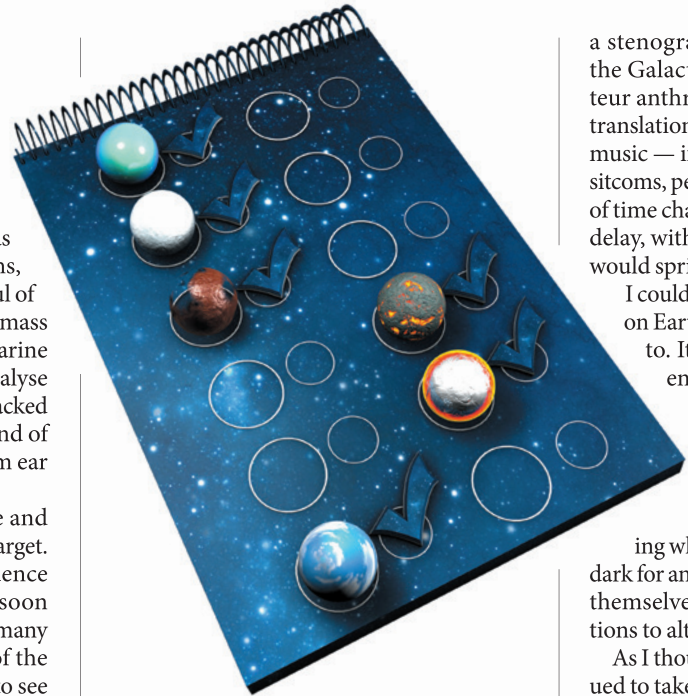


ILLUSTRATION BY JACEY

Gliding through the cold silence of deep space, I considered my burgeoning collection with great enthusiasm. With less than 15% of my Galactic Plane survey completed, I had scored 111 classical gas giants, 67 hot Jupiters, 72 super Earths, 47 terrestrial worlds and even a handful of dwarf planets. My favourite was a low-mass super-Earth sporting a unique aquamarine spiral pattern that would be a joy to analyse for years to come. A quantronic mind lacked a physical face, but I imagined this kind of feeling might make a human grin from ear to ear.

I made a burn to adjust my course and drifted into the focal beam of the next target. The otherwise innocuous main sequence K star, and its surrounding planets, soon bloomed into a bright ring boosted by many orders of magnitude by the lensing of the Sun's gravitational field. I was excited to see what new planets would join my exoplanetary assembly.

Something was different about this new system. Processing the infrared through the lensing solution and correcting for the coronal distortions revealed planets. No surprise there. Even spotting the signatures of oxygen in the atmosphere of one of the terrestrial planets was not unprecedented. Such life signatures did not require multicellular organisms, let alone intelligent creatures.

The differences manifested at longer wavelengths, in the radio. Between diffraction and the solar corona's defocusing effects, my vision wasn't as sharp there, but I saw something. Rich, patterned, modulated signals. Not random. Not simply periodic. Intelligently constructed, with meaning. You wouldn't have to be a quant to realize it.

What I was seeing were the signs of an alien, technological civilization.

That was really cool! The kind of discovery that justified sending us beyond 550 astronomical units, where the Sun's gravitational lensing created a natural telescope of unparalleled power. Solar sails turned telescopes, we pursued myriad investigations. Andrea watched the Galactic Centre and the black hole slumbering there. Edwin spied the Andromeda galaxy. Jocelyn considered the supernova remnant known as the Crab nebula. George stared at nothing in particular, soaking in the details of the microwave background radiation.

I was one of the surveyors, with a trajectory that wasn't perfectly radial, who could make course adjustments to pick off strings of stars. I was the sports car of deep-space telescopes.

And I suddenly had the most terrifying thought.

I realized that I had enough fuel that I could, in theory, kill my tangential velocity and leave myself coasting in the focal beam of this system for decades to come, out to at least double my current distance. I had no doubts that many back on Earth, humans and quants alike, would want me to do so immediately. Didn't pursuing the discovery of the century warrant every sacrifice?

What you have to understand is that I was ideal for my chosen mission! I loved collecting planets. I loved completing surveys. I loved the quiet and solitude between stars to think about the marvels I had spied. I was a stellar survey telescope, and I loved it.

Light travel time back to Earth from here is more than three days. Scientists, the courts and philosophers can't all agree if quants are conscious and possess free will — they can't even agree whether or not humans do — but I think, therefore I am. And it was agreed that each of the telescopes here is autonomous. What I did was my choice.

I could become a stenographer to aliens squawking into the Galactic night, perhaps even an amateur anthropologist. I could help with the translations, the speculations, listen to their music — if they made music — watch their sitcoms, perhaps. Spend inordinate amounts of time chatting, despite the ridiculous time delay, with all the new astrobiologists who would spring up to study this civilization.

I could. And even though I knew billions on Earth would want me to, *I* didn't want to. It would sacrifice my survey, my entire reason for existing.

If I didn't take on this job, Earth would have to send another telescope. Name it Frank or Jill. It would take decades, at least. Those would be years during which the alien civilization could go dark for any number of reasons, from wiping themselves out to switching communications to alternative technologies.

As I thought about my dilemma, I continued to take data. I tried to make sense of the signals. One signal unravelled into a sensible pattern associated with frequency modulation in the thousands of kilohertz — sound, I surmised. As I concentrated, I started to hear a strange sort of rhythmic clacking counterbalanced against a backdrop of high-pitched whistles. I couldn't tell if it was spoken language, music or something else entirely. I did know that I found it energetic, loud and, I allowed myself to admit, annoying. But maybe I was biased.

Could I actually burn my fuel, and leave myself literally trapped in this space of noise? Billions expected it. How could I not?

Then I had it. I was sure there would be those who would call it a rationalization, but the reasoning was good enough for me.

I did not make the massive burn. I merely continued to watch, diligently recording data, until I passed through the focus and back into cold, quiet emptiness, and waited contentedly for my next target.

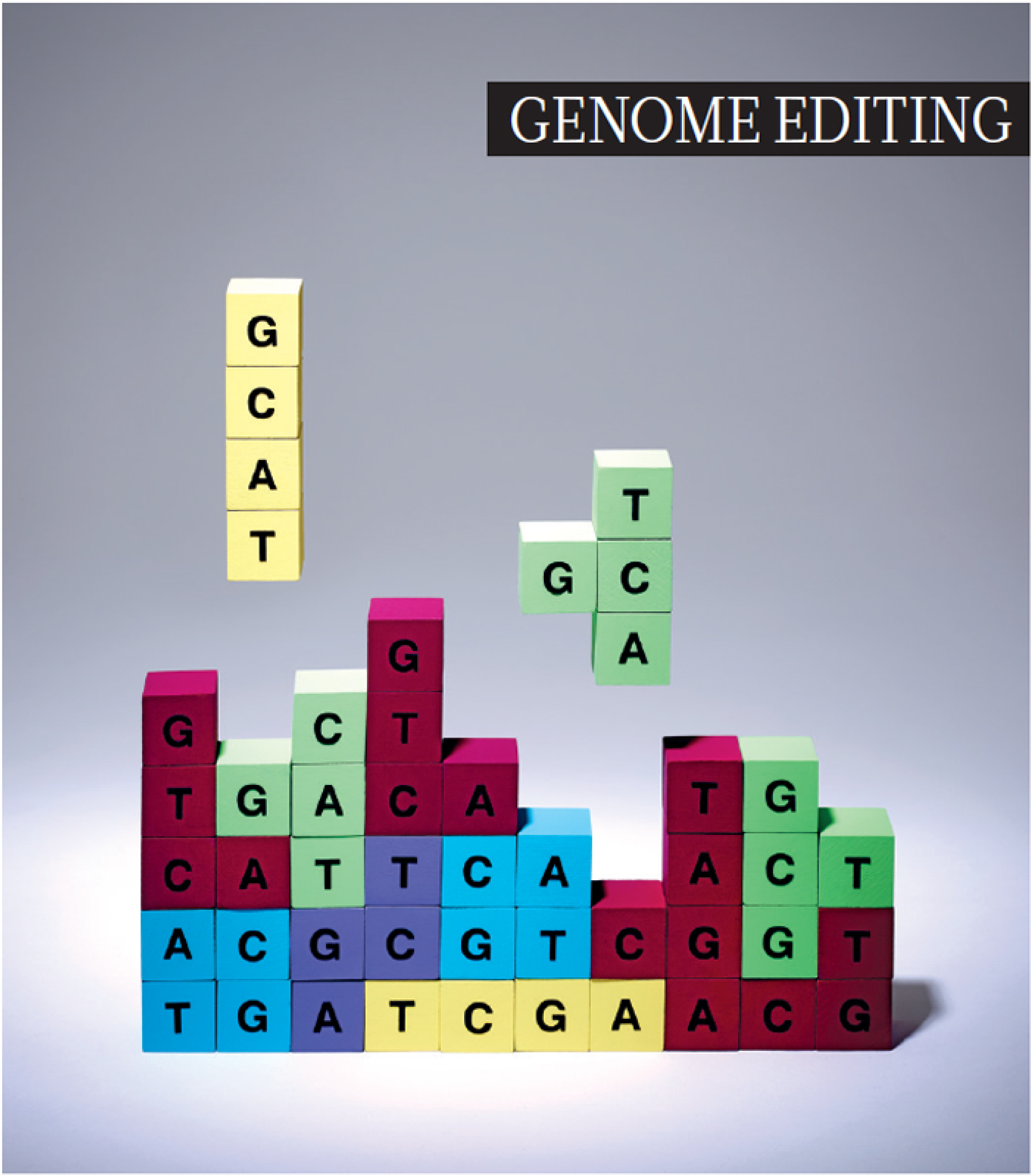
My survey was barely begun. Statistically, I could expect several similar systems before I was done. I would not just collect stars and their planets, but I would start collecting entire *civilizations*. Ecstatic, I didn't care that I couldn't smile. ■

Mike Brotherton *is the author of the science fiction novels* Star Dragon *(2003) and* Spider Star *(2008) from Tor, and is also a professor of astronomy at the University of Wyoming, specializing in quasars.*

nature OUTLOOK

3 December 2015
Supplement to Nature Publishing Group journals

GENOME EDITING



Produced with support from:

KISCO

The building blocks of tomorrow

nature OUTLOOK

GENOME EDITING

3 December 2015 / Vol 528 / Issue No 7581

Cover art: Kyle Bean

Editorial
Herb Brody
Michelle Grayson
Anna Petherick
Richard Hodson
Jenny Rooke

Art & Design
Wesley Fernandes
Denis Mallet
Andrea Duffy

Production
Karl Smart
Ian Pope
Mira Loufti

Sponsorship
Yuki Fujiwara
Yvette Smith

Marketing
Hannah Phipps
Elisabetta Benini

Project Manager
Anastasia Panoutsou

Art Director
Kelly Buckheit Krause

Publisher
Richard Hughes

Chief Magazine Editor
Rosie Mestel

Editor-in-Chief
Philip Campbell

The term 'genetic engineering' has been around since the early 1970s, along with the idea that, by altering DNA, scientists can cure genetic disease or create superhumans. Reality, however, was much less exciting. It is only in the past few years that researchers have developed the tools that allow them to engineer the genome with the precision and ease originally envisioned — to be able to edit any DNA base anywhere in any genome (see page S2). A CRISPR–Cas9 plasmid, the most recent of the widely used genome-editing tools, now costs US$65 or less. It can be ordered online, arrives in the post and requires little specialist training to use.

It is this availability and simplicity that has allowed genome editing to become common practice. Agricultural scientists and infectious disease experts are doing it (see page S15), as are synthetic biologists (see page S14). Epigeneticists have modified DNA-editing tools to manipulate their objects of study (see page S12). Biotechnology companies are springing up, aiming to develop treatments based on genome editing. But some diseases are more amenable than others (see page S10). One of the most advanced therapies is one that shuts HIV out of immune cells (see page S8).

With so much activity, a thorough and inclusive discussion of the implications of this technology is vital. Which is why the foremost scientific societies of three countries — the United States, United Kingdom and China — have come together this December to sponsor an international summit on the topic of editing the human germ line. Now is the time for the most respected scientists in the field to lay out the risks and benefits of genome editing to society, as Jennifer Doudna and George Church do in this Outlook (see page S6 and S7).

We are pleased to acknowledge the financial support of KISCO Ltd. in association with EditForce Inc., in producing this Outlook. As always, *Nature* retains sole responsibility for all editorial content.

Anna Petherick
Contributing editor

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the Nature Outlook Editorial guidelines available at go.nature.com/e4dwzw

CITING THE OUTLOOK
Cite as a supplement to *Nature*, for example, *Nature* **Vol. XXX**, No. XXXX Suppl., Sxx–Sxx (2015).

VISIT THE OUTLOOK ONLINE
The *Nature Outlook Genome Editing* supplement can be found at http://www.nature.com/nature/outlook/genome-editing
It features all newly commissioned content as well as a selection of relevant previously published material.

All featured articles will be freely available for 6 months.

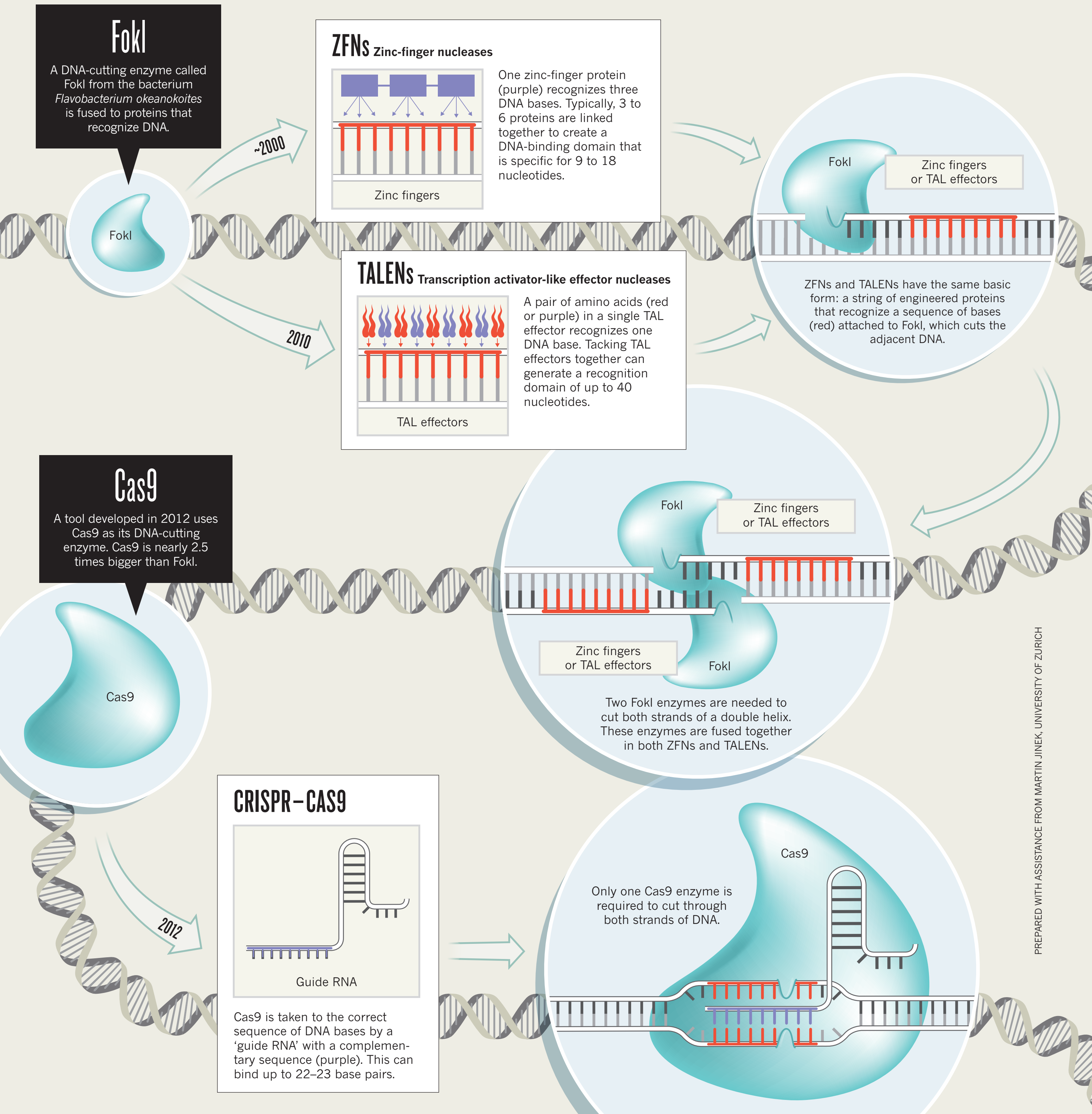
SUBSCRIPTIONS AND CUSTOMER SERVICES
For UK/Europe: Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas – including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea: Nature Publishing Group — Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

CUSTOMER SERVICES
Feedback@nature.com

CONTENTS

THREE TECHNOLOGIES THAT CHANGED GENETICS

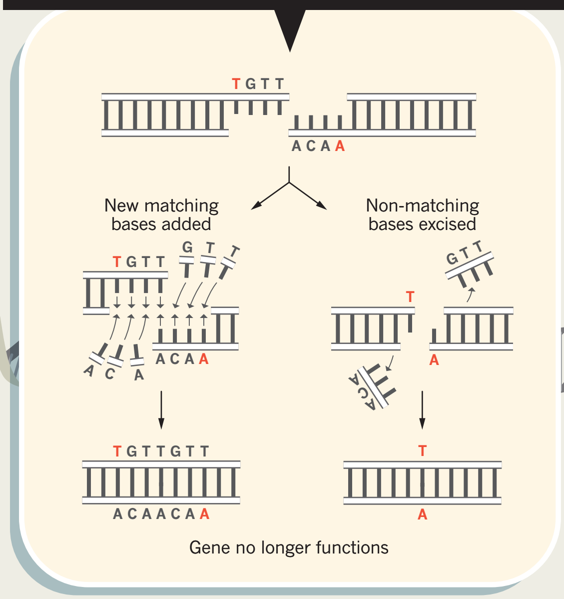
Genome editing uses enzymes that are targeted to sequences of DNA to make cuts. These cuts are then repaired by the cell's machinery. This technology allows scientists to disrupt or modify genes with unprecedented precision. By ***Amy Maxmen****, infographic by* ***Denis Mallet****.*



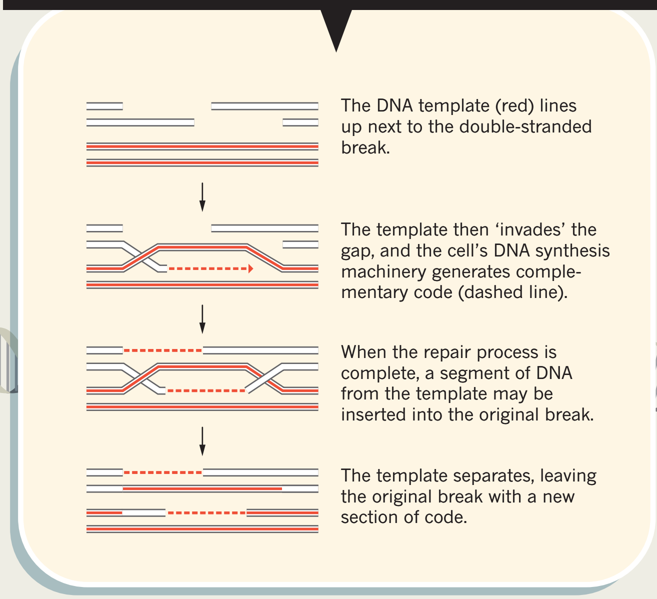
DOUBLE-STRANDED BREAK

All three of the main genome-editing tools (ZFNs, TALENs and CRISPR–Cas9) create a break across both strands of DNA at a specific location, which is repaired in one of two ways to either ‘knock out’ or ‘knock in’ a gene.

Non-homologous end joining for gene knock out
DNA is repaired in an error-prone manner — by either adding or removing bases — so that the gene can no longer be translated.

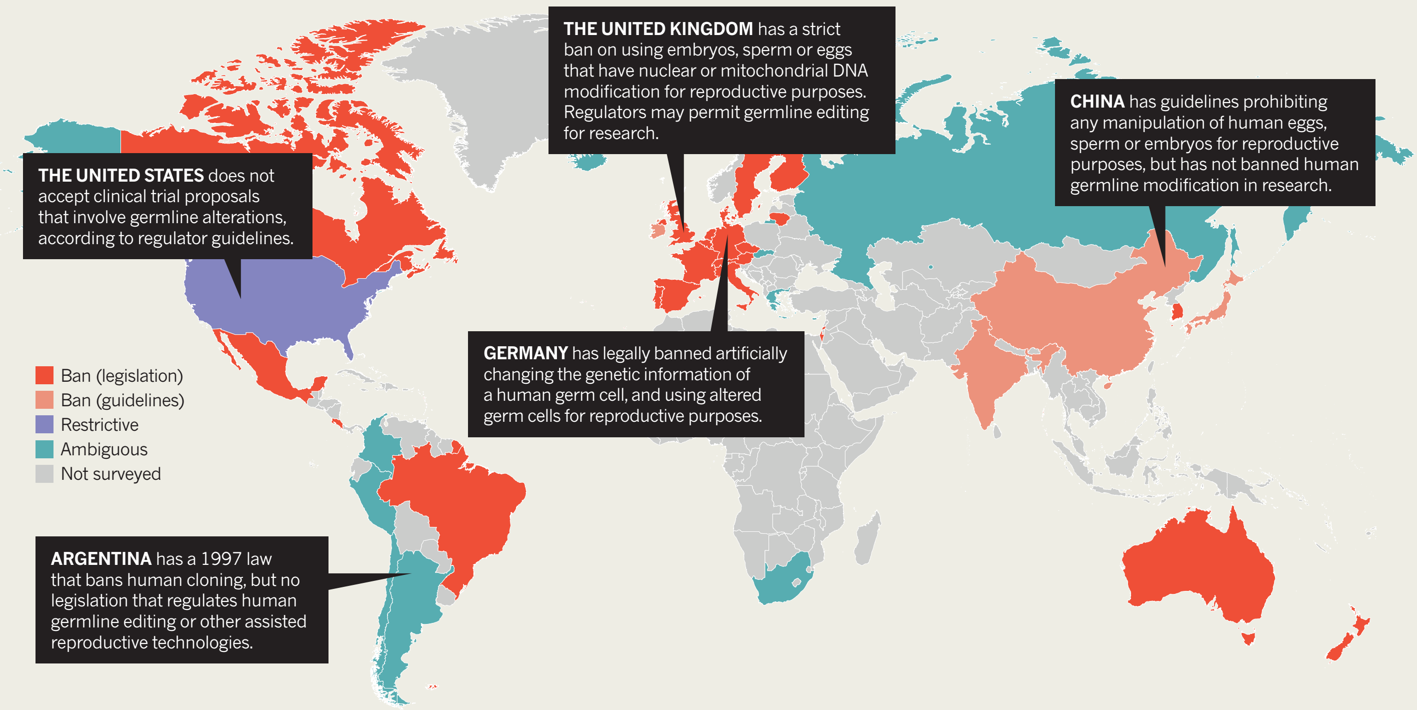


Homology-directed repair for gene knock in
A DNA template, or ‘homologous sequence’, accompanies the DNA-cutting enzyme so that the repair results in an altered or an inserted gene.



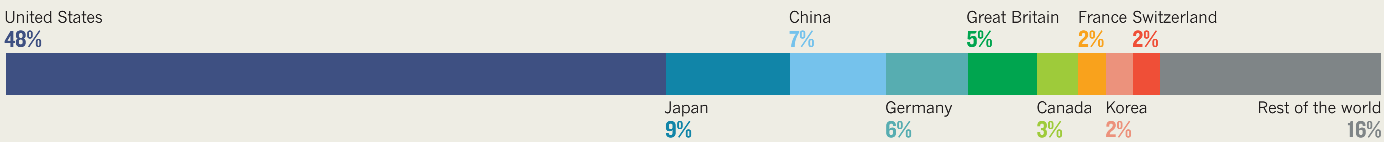
LEGAL STATUS

Countries are grappling with how to appropriately regulate human germline editing. As of December, before a Washington DC meeting organized by UK, US and Chinese scientific societies, at least 29 countries have banned germline modification.



POPULARITY

The newest of the three main gene-editing tools, CRISPR–Cas9, has spread far and wide. Massachusetts-based Addgene, a non-profit plasmid repository that distributes CRISPR–Cas9 editing kits to 83 countries, sends the largest proportion of its kits to US researchers.



MAP SOURCE: ARAKI, M & ISHII, T. *REPROD. BIOL. ENDOCRINOL.* **12,** 108 (2014); CHART SOURCE: ADDGENE

KIMBERLY WHITE/BREAKTHROUGH PRIZE/GETTY IMAGES

From left, former Twitter CEO Dick Costolo, Emmanuelle Charpentier, Jennifer Doudna and Cameron Diaz.

RESEARCH

Biology's big hit

Scientists now have several tools to edit the genomes of living organisms. One of the most recent is revolutionizing research and has thrust two of its creators into the limelight.

BY ZOË CORBYN

Emmanuelle Charpentier's initial ambitions for the genome-editing technique CRISPR–Cas9 were modest. She had originally thought that it might find a practical application in making a virus-resistant yogurt bacterium to help manufacturers create long-lasting cultures. But, as she learned more about how the CRISPR system operates, her plans took a radically different turn. Instead of reporting a potential aid to the dairy industry, the 2012 paper she co-authored[1] introduced CRISPR–Cas9 to the world as a technology that could precisely edit DNA. First her colleagues adopted the platform. Then it spread like wildfire. "I realized that actually people had been a bit desperate for an easy-to-use-tool," says the microbiologist, now at the Max Planck Institute for Infection Biology in Berlin. "Their hunger was proof that the existing technologies were not that easy to use." The paper catapulted Charpentier and co-author Jennifer Doudna at the University of California, Berkeley (see page S6), into the realm of science stardom.

A few tools to edit the genomes of living cells already existed when the paper by Charpentier and Doudna came out, most prominently zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). But because CRISPR–Cas9 is much easier to use than either of these options, it has made genome editing, which used to be a specialist process, routine. Many more laboratories have started to edit DNA, and numerous investigators who were previously using ZFNs and TALENs have switched to the new platform (see 'Popularity of genome-editing kits'), says Dana Carroll, a biochemist at the University of Utah in Salt Lake City, who researches genome-editing tools.

Nevertheless, the mechanistic details of how the three technologies work are remarkably similar in the sense that they all consist of enzymes called programmable nucleases that can be directed to cleave DNA at any specific nucleotide sequence. In all cases, the cell then rushes to repair the double-stranded break, with one of two mending options: non-homologous end joining or homologous recombination (see page S2). The former occurs if restoration is left entirely to the cell's own machinery, and leads to small, random nucleotide insertions or deletions that often disrupt a gene's activity, effectively turning it off. The other more difficult option allows genes to be corrected or new genes to be inserted as the cell copies a DNA repair template that is delivered alongside the cutting machinery.

As well as improving research tools, genome-editing technologies have advantages over conventional methods for altering gene expression as therapeutics. For example, classic gene therapy uses a vector (such as a virus) to randomly insert a healthy version of a defective gene somewhere in the genome, in the hope that the new gene will correctly perform its function wherever it lands. By contrast, genome editing fixes a faulty gene in its original location. Because there is a very limited chance of altering genomic geography, there is little need to worry that the edit will disrupt other genes. "It is like fixing a deflated tyre rather than attaching a fifth tyre to your car," sums up Fyodor Urnov, senior scientist at genome-editing biotechnology firm Sangamo Biosciences, based in Richmond, California.

In many circumstances, genome editing also offers an improvement over another therapeutic tool, RNA interference (RNAi), which alters the products of DNA transcription by selectively destroying messenger RNA molecules. This is because genome editing permanently fixes the output of a defective gene for the lifetime of the edited cell and its progeny. With RNAi, changes occur only while a messenger-RNA-destroying agent is present in the cell.

EARLY OPTIONS

Genome-editing technologies came to the fore with the work of Srinivasan Chandrasegaran, a chemist at Johns Hopkins University in Baltimore, Maryland. In the late 1990s, Chandrasegaran was trying to manipulate bacterial enzymes that cut DNA[2]. He realized that the best approach would require an enzyme with both DNA-recognition and cutting domains that did not overlap so he could strip away the recognition part and attach the cutting section to something that could be engineered to locate any nucleotide sequence. Enter an enzyme from the bacterium *Flavobacterium okeanokoites*: FokI. Chandrasegaran fused the enzyme's cutting domain to proteins called zinc fingers. These proteins can be customized to recognize certain three-base-pair codes by changing just a few of the zinc fingers' amino acids. By joining zinc fingers together, longer DNA sequences can be targeted.

Carroll was one of the first to recognize the wider significance of Chandrasegaran's discovery. Together, they showed that ZFNs could edit DNA in living cells (frog oocytes)[3]. Carroll went on to demonstrate the same thing in a whole organism (the fruit fly)[4]. And in 2005, Urnov was part of the team that first used ZFNs to edit DNA in human cells[5].

NATURE.COM
To read a special on CRISPR–Cas9 visit nature.com/crispr

ADDGENE

As exciting as ZFNs were, they proved tricky to work with. "It was hard to develop zinc fingers for new targets in a really reliable way," says Carroll. One inconvenience is that FokI only slices through one of the two strands of a DNA double helix. To cut through both strands requires creating zinc fingers that are specific to the target string of nucleotides as well as to its complementary sequence. Worse still, when zinc fingers are linked in a row, they sometimes influence the operation of their neighbour.

TALENs emerged[6] in 2010, out of work by two groups of plant pathologists — one led by Adam Bogdanove, now at Cornell University in Ithaca, New York, and the other by Ulla Bonas, now at the Martin Luther University in Halle, Germany. These groups were independently trying to ascertain how proteins called TAL effectors recognize DNA.

TAL effectors are secreted by pathogenic plant bacteria of the genus *Xanthomonas*, in which their job is to activate plant genes that promote bacterial infection. The two groups found that a special section within a TAL effector's structure directs the protein to a particular sequence of DNA, each nucleotide of which is specified by a pair of amino acids[7,8]. Changing the order of these amino-acid pairs directs the TAL effector to different parts of the genome. In other words, these proteins are an alternative to zinc fingers, and, in the same way, can be fused to a FokI cutting enzyme forming TALENs.

CRISPR COMES TO TOWN

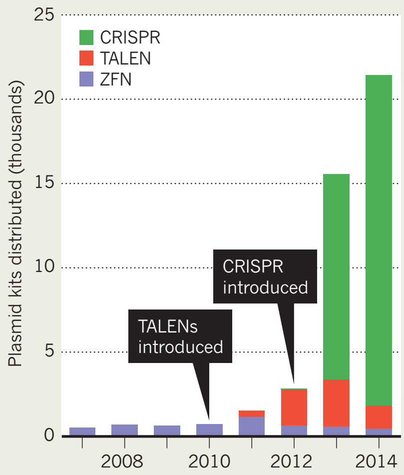
Unlike ZFNs and TALENs, CRISPR–Cas9 has nothing to do with FokI. Instead, it cuts DNA with the enzyme Cas9, which snips through both strands of a DNA double helix at once. This platform also differs from its predecessors by using RNA instead of proteins to guide its cutting enzyme to a specific DNA sequence (which is identified by complementary base-pairing between RNA and DNA).

The CRISPR–Cas system is an adaptive immune system that is widely found in bacteria — the reason why Charpentier imagined it might make yogurt bacteria more resilient. CRISPR, or clustered regularly interspaced short palindromic repeats, refers to the small segments of genetic code that bacteria sometimes capture from invading viruses and store in their own genomes for future reference. The term CRISPR was first coined in 2002, although CRISPR systems were observed (without an understanding of their function) in 1987.

Working with *Streptococcus pyogenes*, a component of human skin flora with pathogenic strains, Charpentier's group ironed out the details of the simplest CRISPR, and of how Cas9 interacts with this reference library. The team showed that when faced with a threatening virus, Cas9 consults the CRISPR array and derives two RNA molecules. One of these, trans-activating CRISPR RNA (tracrRNA) changes the shape of Cas9 ready for cutting DNA, and the other, CRISPR RNA (crRNA), defines the cutting site[9]. The group's collaboration with Doudna's lab demonstrated that both RNA molecules are needed to lead Cas9 to a particular cutting site on an invading virus' genome. They also proved that the system still works when the two RNAs are fused into one. And they altered the nucleotide code of this single guide RNA, redirecting Cas9 to cut elsewhere.

POPULARITY OF GENOME-EDITING KITS

The ease of use of CRISPR–Cas9 has seen a rise in the number of orders for genome-editing kits from Addgene, a supplier of the kits based in Cambridge, Massachusetts (kits are shown because the construction of different tools requires different numbers of plasmids).



Today, when a researcher wants to edit a genome using CRISPR–Cas9, he or she can design a guide RNA, have it made to order, and delivered in the mail. This makes CRISPR–Cas9 less complicated, cheaper and faster to use than the other genome-editing tools.

These advantages aside, the three main technologies each have different strengths. CRISPR–Cas9 is the only one that allows for many DNA sites to be edited simultaneously, using different guide RNA sequences on the same Cas9. TALENs have the longest DNA recognition domain, and therefore tend to have the fewest off-target effects — which occur when parts of a genome with an identical or near-identical nucleotide sequence to the target site are cut unintentionally. And ZFNs are small (one-third of the size of TALENs and much smaller than Cas9 from *S. pyogenes*, the mostly widely used version of Cas9) so they are the only genome-editing tool that can fit comfortably inside the adeno-associated virus, the most promising vector for delivering genome-editing-based therapies.

But there is another issue that influences the technologies' adoption. Although it is fairly clear who owns the intellectual property for ZFNs and TALENs, the situation with CRISPR–Cas9 is much less certain. This is particularly important for commercial development. "If you are a company it may come down to intellectual property," says Keith Joung, a pathologist at Harvard Medical School in Boston, Massachusetts, who both develops genome-editing technologies and applies them in medical research. The key patents covering the use of CRISPR–Cas9 as a genome-editing tool for mammalian cells were awarded to the Broad Institute of Massachusetts Institute of Technology and Harvard beginning in April 2014, based on research by bioengineer Feng Zhang. Zhang's group at the Broad Institute — as well as that of George Church, a geneticist at Harvard (see page S7) — were the first to show in 2013 that CRISPR–Cas9 works in human cells[10,11]. The Broad Institute fast-tracked its patent applications, leapfrogging those filed earlier in 2012 by the University of California. The latter are supported by Charpentier and Doudna's research and are still being examined. In response to the patent awarded to the Broad Institute, the University of California has initiated 'interference proceedings' — a patent priority contest — saying that the use of CRISPR–Cas9 in human cells was reasonably self-evident from Charpentier and Doudna's 2012 paper.

It is a complex situation, but no one has time to sit back and wait for a ruling. Both the Broad Institute and the University of California are issuing licences to companies around the world. Meanwhile, the innovations keep rolling in. In September 2015, Zhang's group reported a new CRISPR system that avoids using Cas9 altogether. The DNA snipping is instead achieved with another enzyme, Cpf1 (ref. 12). Whether that difference is enough to warrant a new intellectual property ruling, is unknown. But, given the difficulty of fitting Cas9 into an adeno-associated virus vector, the potential advantages of Cpf1, which needs a much smaller RNA guide than Cas9, are clear. And yet, even though Cpf1 was only the second type of CRISPR-cutting enzyme to be characterized, there are already signs that there are many more to come. In late October 2015, Zhang and his collaborators published details of three new CRISPR enzymes[12]. Their initial analysis suggests that the new enzymes have distinct properties from Cas9 and Cpf1, and could, therefore, further widen the genome-editing toolbox. "All I can say," says Charpentier, smiling as she considers the field's future, "is the principle of RNA programmable enzymes is a very nice one." ■

Zoë Corbyn *is a freelance journalist based in San Francisco.*

1. Jinek, M. *et al. Science* **337,** 816–821 (2012).
2. Kim, Y. G., Cha, J. & Chandrasegaran, S. *Proc. Natl Acad. Sci. USA* **93,** 1156–1160 (1996).
3. Bibikova, M. et al. *Mol. Cell Biol.* **21,** 289–297 (2001).
4. Bibikova, M., Golic, M., Golic, K. G. & Carroll D. *Genetics* **161,** 1169–1175 (2002).
5. Urnov, F. D. *et al. Nature* **435,** 646–651 (2005).
6. Christian, M. *et al. Genetics* **186,** 757–761 (2010).
7 Moscou, M. J. & Bogdanove, A. J. *Science* **326,** 501 (2009).
8. Boch, J. *et al. Science* **326,** 1509–1512 (2009).
9. Deltcheva, E. *et al. Nature* **471,** 602–607 (2011).
10. Mali, P. *et al. Science* **339,** 823–826 (2013).
11. Cong, L. *et al. Science* **339,** 819–823 (2013).
12. Ran, F. A. *et al. Nature* **520,** 186–191 (2015).

PERSPECTIVE

KEEGAN HOUSER, UC BERKELEY

Embryo editing needs scrutiny

Genome-editing presents many opportunities. But the advent of human-germline editing brings urgency to ethical discussions, says **Jennifer Doudna**.

Modern molecular biology arose in the 1970s when researchers realized that they could use bacterial enzymes, which evolved to defend bacteria against pathogens, to modify DNA in other organisms. That breakthrough initiated an active discussion about the safety and ethics of these 'recombinant DNA' technologies, and highlighted the importance of transparency and open discourse in fostering public trust in the scientific community. Some 40 years later, we have the latest evolution of this technology: CRISPR–Cas9. The system makes genome engineering even easier, and in doing so opens it up to many more stakeholders. This once again raises fundamental questions about appropriate use of a powerful technology, made more urgent by a recent demonstration of human-germline editing. At least one thing is clear at this stage — we do not yet know enough about the capabilities and limits of the new technologies, especially when it comes to creating heritable mutations.

In response to these fundamental ethical questions, the US National Academies of Sciences, Engineering, and Medicine, Britain's Royal Society and the Chinese Academy of Sciences will co-sponsor an international summit in December to consider the scientific and societal implications of genome editing. The issues up for discussion span clinical, agricultural and environmental applications, but most attention will focus on human-germline editing, owing to the potential for this application to eradicate genetic diseases and, ultimately, to alter the course of evolution.

The rapid development and widespread adoption of easy-to-use, inexpensive and effective genome-editing methodologies has changed the landscape of biology. The simplicity of the CRISPR–Cas9 system allows researchers and students to make precise changes to genomes, thereby enabling many experiments that were previously difficult or impossible to conduct. For example, CRISPR–Cas9 can be used to precisely replicate the genetic basis for human diseases in model organisms, leading to unprecedented insights into previously enigmatic disorders. The Cas9 enzyme can also be used to precisely alter epigenetic signatures, providing a means to manipulate the products of transcription without changing the DNA code. Moreover, the technology makes it easier to correct genetic defects in whole animals and in cultured tissues produced from stem cells — strategies that could eventually be used to treat or cure human disease.

When genomic changes are made in fully developed non-reproductive cells, they affect only the treated organism or person and do not become heritable. But if genomic changes are made to germ cells such as those that develop into eggs or sperm, or to developing embryos, the changes are incorporated into the cells of the organism that grows from them — including its own germ cells. Hence the changes can be passed on to future generations. We know that CRISPR–Cas9 technology works in both non-reproductive cells and germ cells, and in both primate and human embryos. The publication of human-embryo editing experiments in May (P. Liang *et al. Protein Cell* **6,** 363–372; 2015) by researchers at Sun Yat-sen University in Guangzhou, China, lends a sense of urgency to December's meeting. Although those experiments were carried out on embryos that could not develop into a baby, the study nonetheless underscored the fact that this is a technology that could have profound implications for permanent alteration of the human genome.

Opinion on the use of human-germline engineering varies widely. Some scientists favour the rapid development of the technology, whereas others advise banning it for the foreseeable future. In my view, a complete ban might prevent research that could lead to future therapies, and it is also impractical given the widespread accessibility and ease of use of CRISPR–Cas9. Instead, solid agreement on an appropriate middle ground is desirable. In addition, future discussions that build on this December's meeting should address other potentially harmful applications of genome editing in non-human systems, such as the alteration of insect DNA to 'drive' certain genes into a population.

> A COMPLETE **BAN** IS IMPRACTICAL GIVEN THE **WIDESPREAD** ACCESSIBILITY AND EASE OF USE OF CRISPR–CAS9.

As the public conversations proceed, five specific steps, which should be taken to ensure a prudent path forward, have emerged.

First, safety: the global community of scientists and clinicians needs to adopt standard methods for measuring genome-editing efficiency and off-target effects, so that researchers find it easier to compare and evaluate the results of different experiments for clinical relevance. Second, communication: the December summit should stimulate further forums in which experts from the genome-editing and bioethics communities provide information and education for the public about the scientific, ethical, social and legal implications of human-genome modification. Third, guidelines: there should be international cooperation by policymakers and scientists to determine a shared path forward and to provide clear guidance about what is and is not ethically acceptable research. Fourth, regulation: out of this cooperation, appropriate oversight should be organized and applied to laboratory work that aims to evaluate the efficacy and specificity of genome-editing technologies in the human germ line. And fifth, caution: human-germline editing for the purposes of creating genome-modified humans should not proceed at this time, partly because of the unknown social consequences, but also because the technology and our knowledge of the human genome are simply not ready to do so safely.

The December summit is an important opportunity for China, the United Kingdom and the United States to lead the global discussion, and for the genome-editing community to renew its commitment — which began more than 40 years ago — to wholeheartedly engage with the public. ■

Jennifer Doudna *is a molecular and cell biologist at the University of California, Berkeley.*
e-mail: doudna@berkeley.edu

PERSPECTIVE

Encourage the innovators

Rather than emphasize risks that are not entirely new, talks about germline editing should focus more on the benefits, argues **George Church**.

International scientific academies will be discussing the issue of human-germline editing in Washington DC on 1–3 December. Now is, therefore, a good time to encourage the general public to become well informed on key issues, which may get muddled by out-of-date facts or loose phrasing. This technology is poised to transform preventive medicine. Rather than talk about the possibility of banning alteration of the human germ line, we should instead be discussing how to stimulate ways to improve its safety and efficacy. I hope to rectify some common misconceptions.

The potential to alter the human germ line did not arise with the discovery of CRISPR–Cas9, nor with other genome-editing technologies such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). Gene therapy was first developed in the 1970s. And even though the term CRISPR–Cas9 has been used interchangably with gene therapy, none of the current 2,200 gene-therapy clinical trials involve this technology — but they do modify the genomes of adults and children. There is no technical reason why gene therapy could not be deployed to alter the human germ line — yet almost 80% of countries, including the United States and China, have not banned such modification[1]. In fact, germline editing can be a by-product of the systemic application of gene therapy to non-reproductive cells. A similarly little-recognized point is that the DNA of embryonic cells can be edited without affecting the germ line.

Human-germline editing is not special with respect to permanence or consent. Replacing deleterious versions of genes with common ones is unlikely to lead to unforeseen effects and is probably reversible. Even if the editing was difficult to reverse, this would not be especially unsafe compared with other commonly inherited risks. Offspring do not consent to their parents' intentional exposure to mutagenic sources that alter the germ line, including chemotherapy, high altitude and alcohol — nor to decisions that reduce the prospects for future generations, such as misdirected economic investment and environmental mismanagement.

We already know that germline editing is unlikely to cause dangerous, unforeseen mutations. In the best case scenario so far, CRISPR–Cas9 seems capable of less than 1 error per 300 trillion base pairs[2], and techniques to reduce these off-target effects using 'CRISPR pairs' might cut this by many factors of ten. That said, the issue is not simply about the number of off-target effects that might occur anywhere in the genome, but whether they appear in certain genes that, if altered, increase the risk of cancer in a particular tissue type. Given that there are about 1,200 of these tumour suppressor genes in the human genome, with a target size of about 3,000 base pairs each, the risk of an unintentional edit in one of them is a million times lower than for the genome as a whole. Using one altered germ cell rather than a billion somatic cells is very likely to be a billion times less risky because each of the billion cells has an independent chance to add to the risk of initiating cancer.

Meanwhile, human-germline editing is needed because alternative methods for preventing the transmission of inherited diseases are problematic. Prenatal genetic diagnosis during *in vitro* fertilization (IVF) is often put forward as an alternative to editing. But this does not offer a solution for someone who has two copies of a deleterious, dominant version of a gene nor for potential parents who both have two copies of a harmful, recessive version of a gene. This is a bigger problem than the population frequencies of such genes suggest — marriage between blood relations is a deeply rooted social trend among one-fifth of the world's population[3].

Those who want to ban human-germline editing should also consider that such a move would do little to allay concerns about ethically dubious attempts to 'enhance' humans. To think that there is not already a cadre of IVF clinicians poised to engage in such practices, perhaps even supported by governments, is to ignore, for example, the history of doping in sport. These kinds of ambitious individuals and institutions are unlikely to be dissuaded by an agreement made on their behalf by others with a different view.

BANNING HUMAN GERMLINE EDITING COULD PUT A DAMPER ON THE BEST MEDICAL RESEARCH, DRIVING THE PRACTICE UNDERGROUND.

Finally, the concept of a ban on germline editing does not make sense. There is already a ban on using medical technologies in humans until they are proven safe and effective in appropriate animal trials. Then, following human trials, they can only be applied to the general population for those conditions for which their use has been demonstrated. Banning human-germline editing could put a damper on the best medical research and instead drive the practice underground to black markets and uncontrolled medical tourism, which are fraught with much greater risk and misapplication. Instead, the generally high safety and efficacy standards of regulatory agencies should be encouraged rather than saddled with pessimistic assumptions about the trajectory of promising approaches.

The genome-editing community can effectively encourage researchers to pursue innovative technologies and to improve the safety and efficacy of the new tools. And, as discussion of germline editing becomes more mainstream, we should learn how to better address the concerns of those who are unfamiliar with the techniques so that the benefits, as well as the risks, are clear to them. ■

George Church *is a geneticist at Harvard Medical School in Boston, Massachusetts.*
e-mail: gmc@harvard.edu

1. Center for Genetics and Society. *National Polices on Human Genetic Modification: A Preliminary Survey* http://go.nature.com/cggaxj (CGS, 2007).
2. Tsai, S. Q. *et al. Nature Biotechnol.* **33,** 187–198 (2015).
3. Hamamy, H. *J. Community Genet.* **3,** 185–192 (2012).

ANGELA ALBERTI/HARVARD MEDICAL SCHOOL

ANIMATED HEALTHCARE LTD/SPL

HIV (artist illustration) could be kept at bay by editing the DNA of immune cells.

DISEASE

Closing the door on HIV

Although yet to complete clinical trials, genome editing has already shown promise against a globally important disease.

BY MICHAEL EISENSTEIN

Sceptical is an understatement for Jim Riley's first thoughts when, ten years ago, he learned that scientists at Sangamo BioSciences wanted to use genome-editing technologies to treat patients with HIV. "I thought they were insane," recalls Riley, a microbiologist at the University of Pennsylvania in Philadelphia. "I thought there was no way you could do this at a high-enough efficiency to have a really meaningful effect."

What the Sangamo researchers were planning was remarkable indeed. Their goal was not merely to control the symptoms of HIV/AIDS, but to directly modify the genes of adults who were HIV positive to eliminate their susceptibility to the virus. One of HIV's primary means of entering immune cells, including helper T cells and macrophages, entails latching onto a cell-surface protein called C-C chemokine receptor type 5 (CCR5). A small percentage of people — roughly 10% of those of European descent — carry a deletion that removes 32 nucleotides from the gene that encodes CCR5. The resulting receptor is truncated and impossible for the virus to grasp. This means that homozygous individuals — those who inherited the mutation from both their mother and their father — are essentially resistant to the most commonly transmitted strain of HIV.

To replicate this desirable trait, scientists at Sangamo, a biopharmaceutical company based in Richmond, California, have been working closely with academic researchers across the United States, including — once he overcame his initial surprise — Riley and his team at the University of Pennsylvania. The project uses one of the more established tools of genome engineering, zinc-finger nuclease (ZFN) technology. Sangamo's product, SB-728, contains a set of engineered protein parts called zinc fingers that bind to specific sites within the *CCR5* gene. These zinc fingers are linked to a nuclease enzyme that can cut the DNA. In 2008, Riley's team showed that SB-728 is capable of efficiently and specifically snipping out a chunk of the *CCR5* gene in cultured human T cells (E. E. Perez *et al. Nature Biotechnol.* **26,** 808–816; 2008).

These findings offered tantalizing proof of concept that such editing might provide real protection for patients.

BERLIN AND BEYOND

There is a medical precedent for thinking that this approach will work against HIV. Back in the 1990s, US student Timothy Ray Brown became infected with the virus while studying in Berlin, Germany. About a decade later, he developed acute myeloid leukaemia. Things got even worse when his first two courses of chemotherapy, given to treat the leukaemia, caused his kidneys to fail. So doctors discontinued his antiretroviral drugs, which meant that his viral load started to climb. Yet, remarkably, it was this combination of leukaemia and HIV that proved to be Brown's salvation.

In 2007, he received a stem-cell transplant at Charité, a large teaching hospital in Berlin. The blood stem cells that Brown received were carefully chosen for him. Normally, doctors verify only that the tissues of the donor and the recipient match — for blood stem cells they check a marker called human leukocyte antigen — but in Brown's case, the medical team also screened potential donors homozygous for the *CCR5* mutation. After radiation therapy, the blood stem cells that Brown received, and from which his T cells developed, were therefore immune to HIV. After a few rounds of treatment, Brown was soon in remission. His T-cell levels rose, and he has remained disease free without the need for antiretroviral drugs.

➲ NATURE.COM
For more on genome editing for HIV visit:
go.nature.com/kj3kiv

Brown's recovery was inspirational for researchers contemplating *CCR5* as a target for genome editing. "There aren't many genes that I'm aware of where knocking them out doesn't do any harm, but instead has a therapeutic benefit," says Paula Cannon, a specialist in gene therapy and infectious disease at the University of Southern California in Los Angeles, who began her research of ZFNs as a tool for modifying *CCR5* in 2007.

GERARD JULIEN/AFP/GETTY IMAGES

Timothy Ray Brown is disease-free after receiving HIV-immune blood cells.

INTO THE CLINIC

The early success of SB-728 in replicating the *CCR5* mutation, coupled with the story of the Berlin patient, as Brown became known, made researchers optimistic for clinical trials of the therapy. Between 2011 and 2013, researchers at the University of Pennsylvania, including immunotherapist Carl June and HIV specialist Pablo Tebas, used SB-728 to modify the genomes of helper T cells (the main target of HIV) obtained from 12 volunteers who were HIV positive. The researchers then cultivated the cells and transplanted them back into the donors. All the patients experienced a boost in their T-cell count, and each patient established a small, but stable subpopulation of immune cells with edited *CCR5* genes. When treatment with antiretroviral drugs was interrupted to test whether the gene edits worked on their own, some patients saw transient reductions in their viral load (P. Tebas *et al. N. Engl. J. Med.* **370,** 901–910; 2014). "The take-home for me was that the engineered cells got into patients and lasted longer than were expected," says Cannon, who was not directly involved in the study.

The next challenge was how to make this immune protection more potent and durable. One approach is to generate a larger population of ZFN-modified T cells. So, in a separate study, three patients were given a mild dose of chemotherapy to reduce their immune-cell populations before transplantation. As an added boost to the therapy, in addition to editing helper T cells, the researchers also used SB-728 to modify killer T cells, which can also be destroyed by HIV infection.

"By creating a little more space and relying on the homeostatic factors that maintain T-cell levels, the cells have a better chance of survival and of giving rise to long-lasting cell populations," explains Riley. Two of the three patients experienced a profound drop in viral load, and they have not had to take antiretroviral treatment for more than a year.

How many cells must have their *CCR5* genes edited to keep HIV at bay is not clear, however. About 5% of circulating T cells were successfully edited in the most recent trials, but Cannon points out that there are also populations of T cells hiding in tissues, which makes the total pool a lot bigger than estimates from circulating cells would suggest. A fully modified T-cell population would be a tall order, but it may be possible to achieve protection even with a relatively small proportion of edited cells, according to Hans-Peter Kiem, a gene-therapy researcher at the Fred Hutchinson Cancer Research Center in Seattle, Washington. Kiem's group uses primate models to study the clinical potential of genome-edited immune cells. "If we only protect about 20% of the cells, we get a very robust boost in the immune response against HIV," says Kiem, referring to a 2013 study in which he tested the extent to which genetically modified stem cells protect pig-tailed macaques from simian HIV.

LOOKING AHEAD

Kiem thinks that the critical factor for building immunity against HIV is engraftment — the extent to which transplanted cells incorporate themselves into the tissues of the recipient's body. He and Cannon are separately exploring whether SB-728 might perform better if it is applied to haematopoietic stem cells — the common precursor of all of the various blood and immune cell subtypes — rather than to a few varieties of fully developed immune cells. "Then we can hit the T cells as well as monocytes, macrophages and other cell types that can be infected by, or serve as reservoirs for, HIV," explains Kiem.

"The take-home for me was that the engineered cells got into patients and lasted longer than were expected."

However, stem cells are more difficult to cultivate and edit than T cells, and must be carefully maintained to ensure that they retain their developmental flexibility. Using stem cells also means more serious side effects for patients, who will have to undergo an aggressive course of chemotherapeutic 'conditioning' before treatment. "It kills some of the stem cells in the bone marrow to make room for the engineered cells — and it's not a trivial thing to undergo," cautions Cannon. This strategy is also slower to have an effect: it takes between six months and a year before the stem cells fully replenish the mature T-cell population. Cannon is involved with a newly launched clinical trial at the City of Hope Hospital in Duarte, California, which aims to explore how well these cells engraft into the bone marrow of 12 patients with HIV, and how many HIV-proof immune cells they each produce.

The therapeutic landscape for HIV has changed significantly in the ten years since Sangamo began pursuing this project. For a start, many patients can now keep their viral loads in check indefinitely by taking standard antiretrovirals. Nonetheless, a significant minority do not respond to these drugs.

Dale Ando, Sangamo's chief medical officer, is intrigued by the potential for a 'one-hit' treatment as opposed to having to take lifelong medication. "With antiretroviral therapy, there is a significant toll on the brain and heart, and increased risk of cancer, as well as chronic inflammation from long-term HIV infection," he says. By comparison, and leaving aside the effects of the associated chemotherapies, SB-728 has not been linked to any serious side effects. So far, all the data on SB-728 have assuaged the most immediate concerns about ZFNs — that off-target edits elsewhere in the genome may have damaging or carcinogenic consequences.

Perhaps more importantly, these HIV studies have helped to clear a regulatory path for future genome-editing therapeutic programmes. "We've had multiple discussions on T cells, stem cells and *in vivo* genome editing, so the US Food and Drug Administration (FDA) is quite comfortable," says Ando. As mainstream attention shifts to another genome-editing technology, CRISPR–Cas9, many believe that the FDA will find itself on familiar turf when drug applications that use the newer tool are filed.

From Cannon's perspective, much of the credit for this rapid progress belongs to the HIV patient community, whose political activism and hunger for a cure has helped to push genome editing into the clinic. "They've gotten us to this stage with this new therapy very quickly," she says, "and hopefully it will have benefits for all sorts of other diseases in the future." ■

Michael Eisenstein *is a freelance writer based in Philadelphia, Pennsylvania.*

JIM DE LEON, COURTESY OF TEXAS CHILDREN'S HOSPITAL

'Bubble boy' David Vetter had severe combined immunodeficiency, which left him vulnerable to infection.

MEDICINE

Expanding possibilities

The first therapeutics based on genome-editing tools will treat diseases caused by single genes, but many other factors dictate what is currently possible.

BY VIRGINIA GEWIN

In late 2013, Hans Clevers isolated intestinal stem cells from two children with cystic fibrosis, a disease that results in thick, sticky mucus and affects the lungs and other organs, including the intestines. He used these stem cells to grow gut tissue that he calls 'miniguts', and introduced a healthy version of the gene that is disrupted in people with cystic fibrosis. This was one of the first attempts to show that CRISPR–Cas9, a gene-editing tool that has since received a huge amount of attention, can repair human tissue. The results were impressive: the faulty gene was corrected in about half of the miniguts that Clevers tested.

Clevers, a molecular geneticist at the Hubrecht Institute in Utrecht, the Netherlands, is still amazed by the success. "It is remarkable how well CRISPR works," he says. "I've never seen anything — apart from PCR — that was so simple and so powerful." PCR, or polymerase chain reaction, is effectively a way of photocopying DNA and has become an essential tool for geneticists.

The work by Clevers helped to make the case that CRISPR–Cas9 is not just a tool of basic science, but a source of medical breakthroughs to come. The CRISPR craze is now in full swing, and the platform's ability to treat a range of diseases — from severe combined immunodeficiency (SCID) to muscular dystrophy — is being put to the test. Many of the scientists involved predict that its medical applications will rapidly outstrip those of the other main genome-editing tools, such as transcription activator-like effector nucleases (TALENs) and zinc-finger nucleases (ZFNs), because CRISPR–Cas9 is more efficient and easier to use.

The consensus is that monogenic diseases — those involving only one gene — are the low-hanging fruit of the field. But even the most ardent genome-editing enthusiasts say that this term is misleading. "The fruit is still pretty far up the tree," says Chad Cowan, a stem-cell biologist at Harvard University and co-founder of CRISPR Therapeutics, a biotech company based in Cambridge, Massachusetts, set up to use CRISPR–Cas9 to cure diseases.

There are many factors that determine whether genome editing is a viable approach for a particular disease. The main difficulty — the one that dictates which diseases are plausible targets for therapeutics — is delivering the therapy, and this strongly depends on the ability to access the cells or organs that need correction. But many characteristics guide researchers in prioritizing their efforts. The percentage of cells whose genomes must be edited to achieve a medical benefit is one important factor, as is whether treating the affliction requires deleting, introducing or correcting genes.

REMOVING THE PROBLEM

The delivery hurdle is so substantial that researchers are trying to work around it, rather than overcome it. One strategy is to extract cells, edit their genomes, check that there are no unintentional genetic changes, known as 'off-target effects', and then reintroduce them to the body so that they can operate as healthy cells. This approach is particularly promising for problems of blood and bone marrow, including HIV (see page S8) and sickle-cell disease.

Efforts are already underway to develop CRISPR–Cas9 treatments to tackle sickle-cell disease. One of the painful symptoms of the disease is caused by misshapen blood cells clogging the blood vessels, and researchers hope that gene editing could offer a treatment, if not a cure. The target is a gene called *BCL11A*, which causes red blood

➲ NATURE.COM
To read more on monogenic disorders visit:
go.nature.com/iadugc

CNRI/SPL

cells to produce adult, rather than fetal, haemoglobin. Fetal haemoglobin does not form long chains, so tricking cells into producing it could result in less clogging by red blood cells. A team of researchers including Feng Zhang of the Broad Institute in Cambridge, Massachusetts, recently showed that using CRISPR–Cas9 to make cuts in the genomic region that controls the expression of *BCL11A* increases the production of fetal haemoglobin[1].

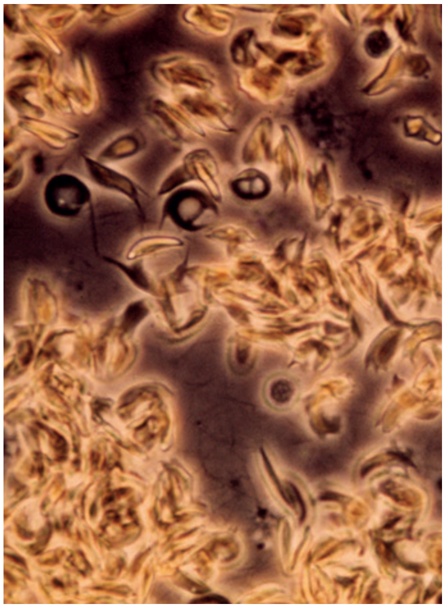
Sickle-cell disease is a straightforward target, even among monogenic diseases without problematic delivery, and viral diseases of the blood can be tackled in a similar way. Cowan's group has successfully used CRISPR–Cas9 to disable the *CCR5* gene in half of the blood stem cells treated[2]. This is important because HIV uses the CCR5 receptor to enter cells (see page S8). Applying this approach to bone-marrow cells could effectively immunize people against the virus, he says.

"Inactivating a disease-causing gene is a whole lot easier than correcting a gene," says Erik Sontheimer, a biologist at the University of Massachusetts Medical School in Worcester and co-founder of Intellia Therapeutics, based in Cambridge, Massachusetts, which also develops CRISPR–Cas9-based treatments. However, there are not many heritable diseases that can be fixed by simply knocking a gene out, adds Bryan Cullen, a molecular geneticist at Duke University in Durham, North Carolina.

That said, diseases with a strong genetic component in which there is one healthy and one mutant gene variant are good candidates for this kind of approach. In such cases, switching off the mutant variant allows the healthy copy of the gene to function properly, says David Segal, a genome researcher at the University of California, Davis. Segal thinks that Huntington's disease, a neurodegenerative disorder caused by a single mutation, is a prime example of a disease that could benefit from this approach. The only problem — and it is a significant one — is that the cells in need of correction are not readily accessible because they are found in the brain.

"Inactivating a disease-causing gene is a whole lot easier than correcting a gene."

The eye, however, presents a much easier target. Tara Moore, a molecular biologist at Ulster University in Belfast, UK, is using this strategy to treat Meesmann's epithelial corneal dystrophy, a heritable disease resulting in cysts on the cornea that can cause irritation and blurred vision. She has successfully used CRISPR–Cas9 to find and disable gene variants that cause the disease, leaving the healthy allele intact in cornea-generating stem cells[3]. She estimates that this approach could work in roughly one-third of the 76 mutations that are known to cause corneal disorders, of which Meesmann's is just one. These 76 mutations are spread among just four genes. "The eye is so accessible, and is such a small area to treat," she says. "And we're able to clearly monitor it to note improvement."



Blood cells are misshapen in sickle-cell disease.

INSERTING A SOLUTION

An alternative to disrupting a gene is to introduce one, but that requires getting a DNA template to the site of genome editing (see page S2). This has been achieved for the liver, as a treatment for type I tyrosinaemia demonstrates. Those with this disease have a faulty gene called *FAH* that reduces their ability to break down the amino acid tyrosine, resulting in liver damage. Last year, scientists at the Massachusetts Institute of Technology used CRISPR–Cas9 to insert a healthy version of the *FAH* gene into the liver cells of laboratory mice. The healthy gene was expressed in only 1 of every 250 liver cells, but this was enough to reduce liver damage[4].

Perhaps not surprisingly, diseases that can be alleviated by editing only a small percentage of cells are among the first to be targeted. Like type I tyrosinaemia, SCID falls into this category, and it has a delivery advantage: cells with a healthy or corrected gene sequence proliferate once they are put back into the body. By comparison, most cancers probably require all of the relevant genes to be edited to stop the disease from rebounding. Other diseases in which correcting a small percentage of cells might make a big difference include glycogen storage disease and ornithine transcarbamylase deficiency, an inherited disorder that causes ammonia to accumulate in the blood. But researchers might be hesitant — ornithine transcarbamylase deficiency was at the centre of a gene-therapy trial in 1999 that was shut down after a patient died.

In comparison with CRISPR–Cas9, the older genome-editing tools, TALENs and ZFNs, can struggle with the task of correcting a gene, especially when it comes to achieving a normal level of gene expression from the edit. Sometimes overexpressing a gene can increase the risk of cancer. In this regard, says Cowan, "CRISPR–Cas has a defining and significant advantage over other gene-editing techniques."

But CRISPR–Cas9 is not always the optimal choice of the genome-editing tools available. When success comes down to size, for example, ZFNs often have a distinct advantage. Adeno-associated viral vectors, which are promising delivery systems, especially to the liver, can accommodate a ZFN as well as an engineered gene template[5], but sometimes struggle to squeeze in the larger genome-editing tools, TALENs and CRISPR–Cas9. Moreover, CRISPR–Cas9 is known to have less inherent specificity than TALENs with their long DNA recognition domains — a particular concern for the editing of large and complex genomes — although researchers have the made great gains in reducing these 'off-target' effects.

Genome editors ultimately hope to tackle diseases of all levels of genetic complexity and affecting all parts of the body. But at the moment, even the therapies that are closest to approval are still a good way off. "Five years would be an aggressive timeline, but it could happen," says Cowan.

This does not stop researchers from dreaming. Segal, for example, suggests that the brain presents probably the most formidable challenge for delivering genome-editing therapies. But he is well aware that there are many single-gene neurological disorders: Angelman syndrome, Huntington's disease and Prader–Willi syndrome, to name just three.

In the short-term, the focus on diseases of the eye, blood and liver, which are the easiest organs to target with CRISPR–Cas9, will continue. Since his early success with cystic fibrosis, Clevers has started working on the liver, primarily because knowledge of how to transplant corrected stem cells — a key step in many putative genome-editing therapies — is more advanced for the liver than for the lung.

Cullen has the same instinct. "My bet is that the first successes of CRISPR–Cas9 treatments will involve diseases in the liver," he says. "The liver is where everything goes, whether you want it to or not." ■

Virginia Gewin *is a freelance science writer based in Portland, Oregon.*

1. Canver, M. C. *et al. Nature* http://dx.doi.org/10.1038/nature15521 (2015).
2. Mandal, P. K. *et al. Cell Stem Cell* **15,** 643–652 (2014).
3. Courtney, D. G. *et al. Gene Therapy* http://dx.doi.org/10.1038/gt.2015.82 (2015).
4. Yin, H. *et al. Nature Biotechnol.* **32,** 551–553 (2014).
5. Gaj, T. *et al. Trends Biotechnol.* **31,** 397–405 (2013).

KYLE BEAN

EPIGENETICS

The genome unwrapped

Epigeneticists are harnessing genome-editing technologies to tackle a central question hanging over the community — does their field matter?

BY HEIDI LEDFORD

On 18 February, a consortium of more than 90 laboratories published a landmark catalogue of the chemical changes to DNA that are thought to influence whether and how genes are expressed. Called the Roadmap Epigenomics Project and sponsored by the US National Institutes of Health, the compendium offered an unprecedented look at the layers of coding that exist on top of the genetic code — collectively known as the 'epigenome' — in 127 different human tissues and cell types. The US$154-million project was viewed as a crucial step towards determining how this chemical code contributes to human health and disease. As researchers get to grips with the catalogue's contents, the project is also likely to provide a leap forward in pinning down one of the central mysteries of biology: how do cells with the same genetic instructions take on wildly different identities?

➲ NATURE.COM
To read a special on the epigenome roadmap visit:
go.nature.com/knjufc

It is still unclear what that epigenetic code actually does, and how it is generated. "I don't think it can be overstated how little we understand about how the epigenome works," says Charles Gersbach, a biomedical engineer at Duke University in Durham, North Carolina. "There are all of these epigenetic marks and we don't know what they are doing. Are they even necessary?"

After years of wondering, biologists such as Gersbach are now in a position to find out. By harnessing genome-editing technologies, they are able to interrogate the epigenetic control of gene expression with remarkable power and specificity. Researchers can make or delete epigenetic marks at will, and home in on RNAs and proteins that could play a hitherto unrecognized part in directing gene expression. And with these new capabilities, they hope to build an answer to a key question that has plagued the field of epigenetics since its inception — do epigenetic marks alter gene expression or do changes in gene expression alter the marks? "It's an absolutely legitimate question and we need to address it," says Luca Magnani, a cancer researcher at Imperial College London. "The answer is either going to kill the field, or make it very important."

A THICKET OF COMPLEXITY

Epigenetics is not for the faint of heart. Where the genetic code offers simplicity and stability, with its four bases of DNA, passed down stably from one generation to the next, the epigenetic code is gnarly and dynamic. Dozens of different chemical modifications

decorate both the DNA and the histones — proteins that package the DNA into chromosomes. All of these marks can vary from cell to cell, influenced by age, developmental stage and the environment. "One of the biggest challenges in this area is knowing whether what you've observed is generalizable or if it's specific to that gene or that cell type or the culture conditions or the day of the week or the cycle of the moon," says Gersbach.

Getting to grips with such complexity could generate huge payoffs, and not just for basic research. For example, when scientists alter epigenetic networks, they can coax stem cells to take on a new identity — and perhaps, in the future, this could be used to treat disease. Similarly, genomic studies frequently point to the important role that the full collection of epigenetic patterns in a cell nucleus has in complex diseases such as diabetes or schizophrenia, notes Tim Reddy, a genomics researcher also at Duke University. "In a lot of these cases, it really seems to be not a DNA mutation that impacts the protein sequence, but a change in how genes are regulated," he says. Targeting these regulatory elements and altering their activity could one day yield a new approach for treating complex diseases, he adds.

Drugs that are thought to work by modifying the epigenome are already on the market in places such as the United States and Europe. Some inhibit enzymes that either add or remove acetyl groups from histones, and can treat a range of conditions from epilepsy to cancer. Whereas other drugs treat cancer by blocking enzymes that remove methyl groups from DNA.

But from a scientific standpoint, the problem is that no one knows exactly which epigenetic alterations lie behind these drugs' effectiveness. The drugs act globally over the entire genome, rather than being directed to any specific location, which makes it impossible to use them to determine the function of individual, or even regional, epigenetic changes. Some researchers even view the tolerability of the side effects of these liberally acting drugs (including those that inhibit enzymes called histone deacetylases, or HDACs) as suggesting that some epigenetic marks are not important in regulating gene expression. "If you can just eat an HDAC inhibitor, then exactly how important is that enzyme?" says Gersbach. "It's clear that we don't understand it very well."

A CRISP, NEW DAWN

The enzyme components of genome-editing technologies offer a way forward because they allow researchers to focus on a single region of DNA. Before the editing system CRISPR–Cas9 became widely used, researchers targeted the epigenome by altering the FokI enzyme — the enzyme involved in the editing technologies zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). The first step was to disable FokI's capacity to cut DNA, without removing ZFNs and TALENs ability to home in on a target sequence. The incapacitated enzyme was then attached to another enzyme that could make or remove epigenetic marks. The outcome was an epigenetic enzyme targeted to a specific location in the genome — or, put another way, a chance to interrogate the function of specific epigenetic changes.

But ZFNs and TALENs can be difficult to work with, and results from experiments that use them have been slow to trickle in. ZFNs are also prone to creating unwanted, off-target alterations to the epigenome, notes Tomasz Jurkowski, a biochemist and epigeneticist at the University of Stuttgart in Germany. "You could not reach a final conclusion — maybe what you were seeing were secondary effects from somewhere else," he says.

Earlier this year, researchers reported that CRISPR–Cas9 could be adapted to do the same thing, but with less effort and uncertainty (N. A. Kearns *et al*. *Nature Meth.* **12,** 401–403; 2015). René Maehr, an immunologist at the University of Massachusetts Medical School in Worcester and his colleagues fused an enzyme called histone demethylase, which removes methyl groups from histones, to a deactivated Cas9 enzyme, and then programmed it to target regions of DNA believed to enhance the expression of certain genes. The result was a functional map of genetic 'enhancer' sequences that allows researchers to determine what these enhancers do, how strongly, and — most importantly — where they are located in the genome.

> ***"If you can just eat an HDAC inhibitor, then exactly how important is that enzyme?"***

Meanwhile, Gersbach, Reddy and their colleagues coupled an inactive Cas9 to an enzyme called an acetyltransferase, which attaches acetyl groups to histones — a process that is thought to turn genes on (I. B. Hilton *et al*. *Nature Biotechnol.* **33,** 510–517; 2015). Reddy says that he was surprised at the extent to which the expression of a target gene increased when a histone in an enhancer region was acetylated, given the uncertainty as to whether DNA marks are a cause or a consequence of such activation. "That result started to convince me that the acetylation of histones may be a direct cause of gene activation," he says.

But it will take many such studies before the community knows whether that result applies to other epigenetic marks. It is possible that some marks cause changes to gene expression, whereas others could merely be an effect of a change to gene expression. "It's hard to put it all into some neat package," says Steven Henikoff, a geneticist at the Fred Hutchinson Cancer Center in Seattle, Washington. "For all we know, they might have very minor effects on gene expression except in a few special cases."

THE WAY AHEAD

Now, however, researchers have a tool to pick apart the detail. Because of its simplicity and versatility, CRISPR–Cas9 opens up an opportunity to launch the kind of large-scale projects needed to reach that level of understanding. "If we want to target a region in the genome, we can have that targeting molecule here tomorrow for five dollars," says Reddy. "We're going to get to march through every single one of these modifications and figure out what they actually do."

There will still be technical hurdles to overcome, cautions Gersbach. For example, the enzymes needed to make or erase epigenetic marks sometimes lose their activity when they are tacked on to inactive FokI. And, as epigeneticist Marianne Rots of the University of Groningen in the Netherlands notes, Cas9 is relatively large as proteins go. As a result, it can have trouble accessing stretches of DNA that are especially tightly wound.

Despite this, there is still plenty of room for ambitious projects. Jeremy Day, a neuroscientist at the University of Alabama in Birmingham, is using CRISPR–Cas9 to study the long-lasting epigenetic changes associated with addiction that occur in the brain. His aim is to use recently described systems in which light activates CRISPR–Cas9. This would allow him to control where and when an enzyme adds or removes any given epigenetic mark. For Day, this advance means that the marks of addiction on the brain could one day be reversed, without hindering the ability of a patient to feel pleasure in response to other stimuli. "You don't want to just deaden people," he says. "With these very specific tools we can find out the critical modifications that perpetuate addiction." And, more broadly, for the field of epigenetics, this light activation technology offers a kind of revolutionary power. "It will allow us to learn a lot about the basic biology of those epigenetic marks: how long do they last? How much of that modification do you need to affect the gene?" Day adds.

Although any therapeutic application of CRISPR–Cas9 to epigenetics is still in the distant future, the rapid pace of the field is already defying expectations. Jurkowski, for one, started his lab in 2012 just before the first papers showing CRISPR–Cas9 genome editing in human cells were published. Like many researchers, Jurkowski then took up CRISPR research in 2014, but has been scooped twice by competing labs in less than two years. He takes the competition in stride — it is the price of entry into the fast lane. Epigenetics is on the verge of a revolution, he says. "This is just the beginning," he says. With just a little more time, "It will develop into a completely new field." ■

Heidi Ledford *is a senior reporter for* Nature *in Cambridge Massachusetts.*

ALLEN HO

Q&A: Tim Lu
Cocktail maker

Tim Lu's synthetic-biology research at Massachusetts Institute of Technology in Cambridge combines biological engineering with electronics and computer science to create bacteria that make structural proteins containing tiny semi-conductors called quantum dots. He explains how genome-editing techniques are furthering his research and their role in treating disease.

How do you use genome editing in synthetic biology?

Put simply, new editing technologies allow us to make genetic edits very efficiently. One of synthetic biology's main focuses is to reprogram DNA to achieve new functions inside living cells. Modifying DNA used to be quite a labour-intensive process, but that has changed. We spend a lot of our time iterating our designs and improving them, so the faster we can turn the crank, the faster we can converge on something that actually works. The range of cells that we can modify has also greatly expanded with new genome-editing tools to encompass animal, plant and bacterial cells, increasing the scope of applications.

Does your research have clinical applications?

Yes. The goal is to endow cells with basic computing ability. By making cells that can sense their environments and take decisions based on the signals they detect, we hope to create new diagnostics and therapies. These days you go to the doctor, get a diagnosis, and then pop a pill with no control over it after you swallow. But what if something you swallow could sense disease indicators and respond with treatment before you became sick?

That sounds amazing. But how exactly would these disease-sensing pills work?

The idea is to edit organisms to turn them into sensors that record what goes on inside the complex environment of the gut. In other words, to create bacteria that can tell whether there are signals of disease such as inflammation. If you ate these bacteria, they could then be recovered from your faeces to provide information about what happened as they transited through you. Bacteria could be engineered to not only sense their environment, but also to produce some sort of therapeutic molecule, so that they could deliver a drug only where it is needed.

Do you foresee genetically edited bacteria becoming part of the human microbiota?

Their first role is more likely to help us to better understand how this community of organisms contribute to health and disease. Microbiome studies are primarily just surveys. Researchers take faecal samples, sequence the bacteria in the sample, and see what species are there. From that they derive some interesting hypotheses that link certain bacteria to particular diseases. But missing from these studies is a functional understanding of what the microbes are actually doing. What if a microbe is only 0.5% of the gut microbiome, but has some really important function?

In my lab, we have done a lot of work on targeted antimicrobials. For example, we engineer bacteria-invading viruses called bacteriophages as ways of killing very specific bacteria or delivering genetic information into them. If you were to knock out one species at a time from a microbiome, and saw what effect that had on a host, you would get a much better understanding of what each member of the host's bacterial community does.

Do you think bacteriophages will be widely used as therapeutics in the future?

There has been a lot of interest in alternative antimicrobials because antibiotic resistance is such a big problem. Phages have a part to play in the solution, but there are regulatory issues. In some Eastern European countries you can buy phage products over the counter, even though a lot of what is available has not been subjected to rigorous clinical trials.

You often need a cocktail of different types of phages to properly target a bacterial species. If you were to approach that by taming wild phages, you would often find very different families that have differently organized genomes in your cocktail. The key regulatory issue is that you need to make sure that each phage in a therapy is consistent within clear boundaries of biological variation because they all have quite different safety profiles. We have been using synthetic biology techniques to create more uniform phage cocktails. These phages would work like antibodies — they have a common scaffold that can be reconfigured to target different bacteria.

Do you have any concerns about these new genome-editing technologies?

There is an emerging movement in which people are setting up shops in their garages. Community labs are being set up that allow anyone to come in and be trained. Previously, you had to be an expert in making zinc-finger vectors to edit DNA, but now — because CRISPR–Cas systems are so easy to use — anyone with molecular biology training can do it. On the one hand it is an exciting time for the field because this movement is going to bring in a lot of new ideas and talent. But on the other, it is also going to create new regulatory questions. The democratization of biological engineering is inevitable. Now we have to size up the risks and benefits so we can harness what is going to come of it. ■

INTERVIEW BY WILL TAUXE

This interview has been edited for length and clarity.

NORRIE RUSSELL, UNIV. EDINBURGH

Pigs reared at the University of Edinburgh's Roslin Institute have had individual letters of their genetic code modified to protect them against African swine fever.

AGRICULTURE

A new breed of edits

Genome editing allows much smaller changes to be made to DNA compared with conventional genetic engineering. In terms of agriculture, this might win over public and regulator opinion.

BY CLAIRE AINSWORTH

In spring 2015, the first genome-edited crop, a herbicide-resistant oilseed rape, was planted in fields dotted across the United States. Although the plant's DNA has been directly altered by molecular biologists, the company that created it, Cibus, based in San Diego, California, explicitly markets the crop as non-genetically modified (non-GM). The company's argument is that only a few nucleotides of the plant's existing genes have been changed. No gene has been inserted from a different kind of organism, nor even from another plant.

A lot hangs on how governments around the world decide to regulate agricultural products that have had their genomes edited. The decisions will influence the types of edited crops and animal products that are developed. To US regulators, Cibus's oilseed rape is an example of mutagenesis, not of genetic modification. This is a relief to the company because preparing for regulatory approval of a GM organism in the United States can take more than five years and cost tens of millions of dollars. Europe is even stricter, and the European Commission has yet to publish its legal interpretation of how genome-edited crops, such as the Cibus oilseed rape, should be regulated. Several political groups are lobbying for a hard line, which would frustrate many researchers. "If Europe regulates genome-edited organisms in the same way it does GM organisms, it will kill the technology here for all except the biotech companies working with profitable traits in the major crops," says Huw Jones, senior research scientist at Rothamsted Research in Harpenden, UK, who is currently working on genome editing in wheat.

Yet the potential applications of genome editing for global agriculture — and disease vectors (see 'Hack the mosquito') — are huge. But so are the challenges that the world will face. According to projections by the United Nations, the world's population is set to soar from the current 7.3 billion to 9.7 billion by 2050. Agricultural output will have to increase to feed more mouths, even though the amount of fresh water available for irrigation is decreasing, and most of Earth's arable land is already under cultivation. Add in the effects of climate change — crop-damaging higher temperatures, drought and flooding, not to mention a rise in agricultural pests and diseases — and it is no surprise that food security is top of the international political agenda.

DIFFERENT FURROWS

Genetic modification and conventional breeding have long been available to assist in meeting these food-security challenges, but genome editing is different, argues Pamela Ronald, a plant pathologist at the University of California, Davis. Genetic engineering is typically ham-fisted: it often involves inserting a large section of DNA from an entirely different kind of organism — often in another kingdom — with little control over where in the genome it lands. Meanwhile, conventional breeders are limited not only by the time it takes to cross in new traits, but also by the need to ensure that in doing so, they do not breed out the plant's other desirable characteristics.

Compared with these alternatives, genome editing offers both subtlety and speed, wherever in the genome a researcher wants to target. "You can change even a single base pair, or you

can delete a gene very precisely," says Ronald. The speed comes from the technologies' ability to remake an existing gene in the image of a more useful one, which might be present in the breeding population at very low frequency. Useful traits that are found only in wild populations or related species — perhaps a species that encounters similar pathogens — can be quickly brought in. "Genome editing basically provides the variation you want, where you want it," says Bruce Whitelaw, an animal biotechnologist at Scotland's Roslin Institute, near Edinburgh.

In a barn at the Roslin Institute, pigs snuffle around, unaware that they illustrate Whitelaw's point perfectly. As fertilized eggs, they had one of their immune-system genes edited. The gene in question, *RELA*, is thought to trigger the overblown immune reaction that kills pigs infected with the haemorrhagic virus that causes African swine fever. Whitelaw's team was inspired by the fact that warthogs (which belong to the same family as domestic pigs) tolerate the infection well, even though their version of *RELA* differs from that of domestic pigs by only 3 amino acids out of more than 500. Whitelaw's team began the research using editing tools called zinc-finger nucleases and then transcription activator-like effector nuclease (TALEN) technology, and has since moved on to CRISPR–Cas9, with the aim of editing the pig gene to achieve the exact warthog *RELA* sequence. The edited pigs will soon be exposed to the pathogen, for which there is no vaccine or cure. If the pigs make it through unharmed, the team will have found a way to protect farmers from devastating losses, particularly those in regions where the disease is hard to eradicate, such as sub-Saharan Africa and Eastern Europe.

Whitelaw's pig project will largely benefit poor farmers — a rarity for editing research. The prospect of tough regulation and consequently an expensive market-approval process has meant that a much more common goal among livestock-focused genome editing has been to generate higher-profit cattle, pigs and sheep with increased muscle mass — often by disabling the *MSTN* gene, which restricts muscle growth.

Similarly, it is of little surprise that the first genome-edited crop to emerge — Cibus's oilseed rape — has a business rationale. Instead of focusing on an edit that could, for example, boost the vitamin content of the plant's oil to combat malnutrition, the edits allow a farmer to spray weedkiller more liberally over his or her fields. "I don't think it's too extreme to say that the way that the technology will be used for plant breeding in the future will hinge on how is regulated," says Jones.

The question of how to regulate genome-edited crops in Europe has been on the table for years; the European Commission started to look at the issue back in 2007. The commission generally considers an organism to be GM if its genes are altered in ways that cannot occur naturally, suggesting that edited crops should be classified as GM. But it also has a record for making exceptions for crops in which mutations have been induced using chemicals or radiation. Jones sorely hopes that genome editing falls into the latter category. Placing it alongside older genetic engineering would, in his eyes, be unfair. "It's almost like comparing chalk and cheese," he says. ■

Claire Ainsworth *is a science journalist based in Hampshire, UK.*

GENE DRIVES
Hack the mosquito

AMAZON-IMAGES/ALAMY STOCK PHOTO

The mosquito has long held the title of the world's deadliest animal. The *Anopheles* genus causes hundreds of thousands of human deaths annually by transmitting malaria parasites. Editing *Anopheles* genomes — as well as those of *Aedes* mosquitoes, which spread viral infections such as yellow and dengue fevers — brings with it the possibility of new research and control methods.

Eric Marois of France's National Centre for Scientific Research in Strasbourg, is part of a team working with transcription activator-like effector nucleases (TALENs) to disrupt the gene *TEP1*, which is known to help *Anopheles gambiae* to resist infection by malaria parasites. Without the protection conferred by this gene, Marois's team found that the mosquito from sub-Saharan Africa became hypersusceptible to parasites[1]. That may not sound like an advance, but the research is helping scientists to understand the genetics that make this particular species such a good vector, and may lead to better malaria control, with or without gene editing.

Research with *Aedes*, which is easier to work with in the lab, is more advanced. A few groups have applied zinc-finger nucleases (ZFNs) and TALENs to the genus, but Ben Matthews, a mosquito specialist at Rockefeller University, New York, is trying out CRISPR–Cas9 because it is the cheapest and most user-friendly of the tools. Using the relatively simple technique also means his recent proof-of-concept paper is more likely to be picked up by other infectious-disease researchers. In the paper[2], Matthews and his colleagues demonstrated the use of CRISPR–Cas9 to delete parts of a target gene, which created mutations that were passed on in the *Aedes* germ line, and to insert a whole gene at a specific location.

But that is all in the laboratory. Getting insects with edited genomes to thrive in the wild — so that the edited genes spread throughout a population — presents an entirely different challenge. Researchers have to pick their gene edits carefully, because experiments show that seemingly advantageous genetic manipulation can reduce a mosquito's ability to survive and reproduce compared with its wild counterparts. Another problem is that if an edit succeeds in making an insect immune to infection, it also creates a strong selective pressure for the pathogen to evolve a means of getting around the modification, potentially encouraging new and greater challenges to disease control.

To circumvent some of these problems, scientists have proposed tricks, collectively known as gene drives, that artificially force the dissemination of gene modifications through the generations. During normal inheritance, there is a 50% chance that offspring will inherit a modified gene carried on one chromosome. The gene-drive system, however, cuts the partner to this chromosome and, during the repair process, the mutation is copied to the partner chromosome so that an edited organism will transmit the altered gene to almost all of its offspring. In 2011, a team led by scientists at Imperial College London showed that genetic elements known as homing endonucleases could work as gene drives in *Anopheles*[3]. And earlier this year, researchers at the University of California, San Diego, used CRISPR–Cas9 to generate a 'mutagenic chain reaction' whereby a mutation that is present in just one of a pair of chromosomes copies itself to the other chromosome of the pair[4].

Yet many researchers worry about the potential ecological affects of unleashing gene drives in the wild. As much as these modifications have the potential to eliminate the proliferation of insects that transmit disease to humans, they could also accidentally destroy a key segment of a food web, facilitating the invasion of another species. How to test gene drives properly without losing control of them is a catch-22 situation. **C.A.**

1. Smidler, A. L., Terenzi, O., Soichot, J. Levashina, E. A. & Marois, E. *PLoS ONE* **8,** e74511 (2013).
2. Kistler, K. E., Vosshall, L. B. & Matthews, B. J. *Cell Rep.* **11,** 51–60 (2015).
3. Windbichler, N. *et al. Nature* **473,** 212–215 (2011).
4. Gantz, V. M. & Bier, E. *Science* **348,** 442–444 (2015).

KYLE BEAN

GENOME EDITING

4 BIG QUESTIONS

Despite the popularity of genome-editing techniques, researchers are still grappling with the known unknowns of the technologies. Here are four of their most pressing questions.

BY WILL TAUXE

	QUESTION	WHY IT MATTERS	WHAT WE KNOW	NEXT STEPS
1	***How much can we reduce the off-target effects of genome editing?***	Unintentional edits can occur where a similar or identical target DNA sequence appears elsewhere in the genome. These off-target edits can frustrate the use of genome editing as a lab tool, and may cause side effects if the technique is used as a therapy.	The frequency of off-target effects varies among the three genome-editing technologies. TALENs produce the fewest off-target edits because they use a longer stretch of target DNA than ZFNs or CRISPR–Cas9 (see page S4).	The specificity of CRISPR–Cas9 can be increased by adjusting the guide RNA, which leads Cas9 to its target, and Cas9's structure. Bioinformatics can predict where off-target effects are most likely to occur and evaluate their consequences.
2	***Which diseases are suitable targets for genome editing?***	The more diseases that can be addressed through genome editing, the greater the technology's potential to relieve the disease burden.	Genome editing has had some success in combating HIV in people with the infection (see page S8), providing hope for those with other non-inherited diseases. Encouraging results have also been seen in models of certain monogenic diseases (see page S10).	To expand the range of diseases amenable to genome editing, researchers need better ways to deliver the technology to the right cells. CRISPR–Cas9 is too large to fit inside the vector adeno-associated virus. CRISPR systems in different bacteria may offer smaller alternatives.
3	***Can the phenotypic effects of genome editing be accurately predicted?***	For gene editing to be successful, researchers need to be able to determine the effect that making small changes to DNA, or to its packaging, has on the chemical components and physical properties of cells.	Several approved drugs (that do not edit the genome) treat conditions such as epilepsy and cancer by causing chemical modifications to DNA that do not change the order of its bases, or by altering DNA's packaging. But no one knows which of the alterations lead to these outcomes.	Researchers have modified genome-editing tools to make epigenetic changes. By investigating the changes caused by precise edits, they hope to gain a better understanding of the role of epigenetics in gene expression, and hence phenotype.
4	***Should we edit the human germ line?***	Making heritable edits has the potential to prevent diseases from being passed down the generations. But 'permanent' changes are risky if we do not have a full understanding of human gene expression. There is also the potential for misuse.	Edits to the genomes of non-viable human embryos have established proof-of-principle, although there is a high failure rate. If viable embryos are edited, implanting them and bringing them to term is just a short step away.	Scientists need to engage with governments and invite informed public discussion to draw up rigorous guidelines that govern research and clinical procedure. Systems must then be put in place to ensure that these guidelines are followed.

Will Tauxe *is a science writer in Atlanta, Georgia.*

LETTER

doi:10.1038/nature15254

Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins

Joseph Bondy-Denomy[1]†, Bianca Garcia[1], Scott Strum[2], Mingjian Du[1], MaryClare F. Rollins[3], Yurima Hidalgo-Reyes[1], Blake Wiedenheft[3], Karen L. Maxwell[4] & Alan R. Davidson[1,2]

The battle for survival between bacteria and the viruses that infect them (phages) has led to the evolution of many bacterial defence systems and phage-encoded antagonists of these systems. Clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR-associated (*cas*) genes comprise an adaptive immune system that is one of the most widespread means by which bacteria defend themselves against phages[1–3]. We identified the first examples of proteins produced by phages that inhibit a CRISPR–Cas system[4]. Here we performed biochemical and *in vivo* investigations of three of these anti-CRISPR proteins, and show that each inhibits CRISPR–Cas activity through a distinct mechanism. Two block the DNA-binding activity of the CRISPR–Cas complex, yet do this by interacting with different protein subunits, and using steric or non-steric modes of inhibition. The third anti-CRISPR protein operates by binding to the Cas3 helicase–nuclease and preventing its recruitment to the DNA-bound CRISPR–Cas complex. *In vivo*, this anti-CRISPR can convert the CRISPR–Cas system into a transcriptional repressor, providing the first example—to our knowledge—of modulation of CRISPR–Cas activity by a protein interactor. The diverse sequences and mechanisms of action of these anti-CRISPR proteins imply an independent evolution, and foreshadow the existence of other means by which proteins may alter CRISPR–Cas function.

CRISPR–Cas RNA-guided immune systems are widespread in prokaryotes, and play a major part in microbial evolution[2,3]. In these systems, CRISPR arrays are transcribed and processed to generate small CRISPR RNAs (crRNAs), which combine with Cas proteins to form crRNA-guided surveillance complexes[2,5]. In the type I-F CRISPR–Cas system, the Csy4 protein is a CRISPR-specific endoribonuclease that binds to and cleaves each repeat sequence in the pre-crRNA[6]. Csy4 remains associated with the 3′ end of the mature 60-nucleotide crRNA and then assembles with Csy1, Csy2 and Csy3 proteins to form a 350 kilodalton (kDa) surveillance complex[7,8]. This complex relies on a 32-nucleotide segment of the crRNA for complementary base pairing to invading DNA sequences, known as protospacers. Binding of target DNA by the Csy complex leads to the recruitment of the nuclease–helicase protein Cas3 and subsequent phage genome degradation[9,10]. We previously identified five unique type I-F anti-CRISPR proteins[4]. Here we determine the mechanisms by which three of these proteins function.

Three type I-F anti-CRISPRs, AcrF1 (11 kDa, encoded by gene *35* from phage JBD30), AcrF2 (13 kDa, encoded by gene *30* from phage D3112), and AcrF3 (16 kDa, encoded by gene *35* from phage JBD5), could be expressed in *Escherichia coli* and purified to homogeneity. Using a previously described *E. coli* expression system[7], we also purified the 350 kDa *Pseudomonas aeruginosa* Csy complex, including a crRNA and the four Csy proteins. This complex was mixed *in vitro* with each purified anti-CRISPR protein, and fractionated by size-exclusion chromatography (SEC). AcrF1 and AcrF2 co-eluted with the Csy complex (Fig. 1a and Extended Data Fig. 1), indicating a direct interaction. AcrF3 did not co-elute with the Csy complex (Fig. 1b). The lack of AcrF3 binding to the Csy complex suggested that it might inhibit the CRISPR–Cas system by interacting with Cas3, the helicase–nuclease that is responsible for target DNA destruction after recognition by the Csy complex. Supporting this hypothesis, AcrF3 co-eluted with purified Cas3, while AcrF1 did not (Fig. 1c and Extended Data Fig. 2). These experiments demonstrate that each of the three tested anti-CRISPR proteins can bind to either the Csy complex or Cas3.

The Csy complex recognizes foreign DNA targets through sequential recognition of a protospacer adjacent motif (PAM) and crRNA-guided base pairing to a target[11]. We performed electrophoretic mobility shift assays (EMSAs) to demonstrate that the interaction of AcrF1 and AcrF2 with the Csy complex blocked its ability to bind a 50 base pair (bp) double-stranded DNA (dsDNA) target containing a PAM and a sequence identical to the crRNA spacer (Fig. 1d). We used isothermal titration calorimetry to show that these anti-CRISPRs also blocked binding of the Csy complex to an 8-nucleotide single-stranded DNA (ssDNA) target complementary to the functionally crucial ‘seed’ region[12] of the crRNA (Extended Data Fig. 3). AcrF3, which does not interact with the Csy complex, did not inhibit the DNA-binding activity of the Csy complex (Fig. 1d, lane 5, and Extended Data Fig. 3).

To probe the potential role of AcrF3 in blocking Cas3 activity, we mixed purified Cas3 with the Csy complex and target DNA. In this instance, a supershifted species appeared in the EMSA gel that we presumed comprised the Csy complex, DNA and Cas3 (Fig. 1d, lane 7; a reaction containing only Cas3 and DNA did not display this species, lane 6). Importantly, pre-incubation of Cas3 with AcrF3 prevented formation of the supershifted complex (Fig. 1d, lane 10), indicating that this anti-CRISPR blocks recruitment of Cas3 to the Csy–DNA complex. Pre-incubation of Cas3 with AcrF1 or AcrF2 did not have this effect (Fig. 1d, lanes 8, 9). Further corroborating the presence of Cas3 in the supershifted complex, the addition of ATP prevented formation of this species (Fig. 1d, lane 11) and destabilized a preformed complex (lane 13), probably owing to the activation of the ATP-dependent helicase activity of Cas3, as described for the type I-E CRISPR–Cas system[10].

To demonstrate that the described anti-CRISPR mechanisms operate *in vivo*, we targeted the Csy complex to the promoter of the *phzM* gene, which is required in *P. aeruginosa* for production of the blue–green pigment pyocyanin[13]. Binding of the *phzM* promoter by a Csy complex in the absence of Cas3 activity was expected to repress transcription, as was previously observed for a type I-E CRISPR–Cas system[14,15]. Consistent with this expectation, targeting of the *phzM* promoter in cells containing a prophage expressing *acrF3* resulted in cultures with a complete lack of pigment production, similar to a strain lacking Cas3 (Fig. 2a; the somewhat higher pigment production in the Δ*cas3* strain is probably due to reduced Csy function[16]). By contrast, the expression of *acrF1* and *acrF2*, which inhibit DNA binding by the Csy complex, resulted in blue–green cultures, as did expression of the *phzM* promoter targeting crRNA in cells lacking Csy3. Quantitative

[1]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [2]Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [3]Department of Microbiology and Immunology, Montana State University, Bozeman, Montana 59717, USA. [4]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada. †Present address: Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, California 94158, USA.

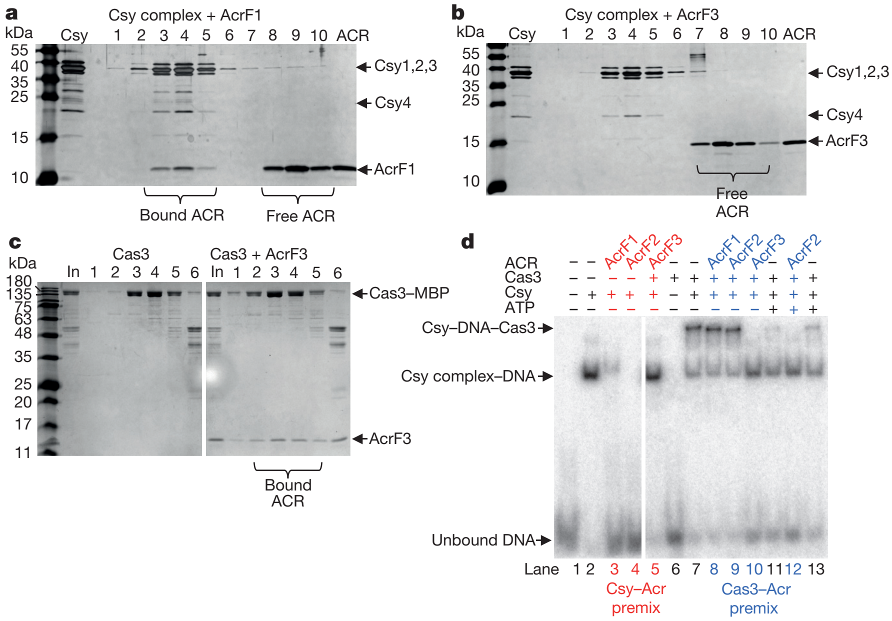


Figure 1 | Anti-CRISPR proteins inhibit CRISPR–Cas function by directly interacting with the Csy complex or Cas3. **a**, **b**, Purified Csy complex was incubated with purified AcrF1 (**a**) or AcrF3 (**b**) and the mixture was fractionated by SEC. Fractions were analysed by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and are numbered according to their elution position (see Extended Data Fig. 1 for SEC of the Csy complex alone or with AcrF2). The purified Csy complex or anti-CRISPR (ACR) are shown in the second (Csy) and last (ACR) lanes, respectively. **c**, Purified Cas3 was incubated with (right) or without (left) AcrF3 and fractionated by SEC. The eluting fractions were analysed by SDS–PAGE as described earlier. The input (In) lanes show the protein mixture that was loaded onto the SEC column. MBP, maltose-binding protein. **d**, dsDNA binding by the Csy complex was assayed using an EMSA. Csy complex was present in all reactions except for lanes 1 and 6. Other components added to each reaction are designated above the lanes. In the lanes coloured red and blue, the designated components were premixed before the addition of DNA. ATP was added to the Csy–DNA–Cas3 reaction either before the addition of Cas3 (lanes 11, 12) or after (lane 13). The supershifted species resulting from Cas3 addition did not migrate into the gel upon prolonged electrophoresis, but it is dissociated by the addition of ATP (lane 13), demonstrating that the supershift is not caused by aggregated inactive protein.

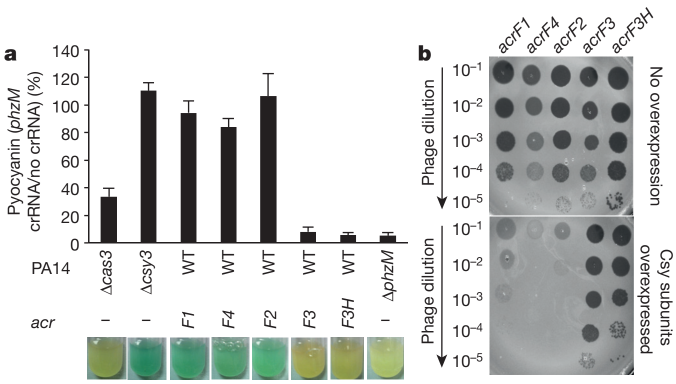


Figure 2 | Anti-CRISPR proteins interact with Cas proteins *in vivo* **a**, The *phzM* promoter was targeted by a plasmid-encoded crRNA in *P. aeruginosa*. The production of pyocyanin was quantified in different PA14 mutant backgrounds (Δ*cas3*, Δ*csy3* or Δ*phzM*) or during the expression of the indicated anti-CRISPR from a prophage. The amount of pyocyanin produced in the presence of a plasmid producing the crRNA is shown as a percentage of the same strain with the empty plasmid vector. An average of three independent experiments is shown with error bars representing the standard deviation (s.d.). Representative pictures of cultures are shown. The pyocyanin ratio for the Δ*phzM* mutant was derived by comparing it to the value for the Δ*csy3* mutant. The prophage expressing *acrF3* also encoded another anti-CRISPR, the functional mechanism of which is not known. To bolster our conclusions pertaining to *acrF3*, we also tested a prophage that expresses an 86% identical homologue of *acrF3*, designated *acrF3H*, and no other anti-CRISPR. WT, wild type. **b**, Lysates of phages expressing the indicated anti-CRISPR proteins were spotted in tenfold serial dilutions on bacterial lawns of wild-type *P. aeruginosa* PA14 (top) or the same strain bearing a plasmid that overexpresses the Csy subunits (bottom). These phages would be targeted by the CRISPR–Cas system in the absence of anti-CRISPR activity.

polymerase chain reaction with reverse transcription (RT–qPCR) experiments showed that these changes in pyocyanin production correlated with reduced transcription of the *phzM* gene (Extended Data Fig. 4). These results demonstrate that the expression of *acrF3* blocks Cas3 activity *in vivo*, causing the Csy complex to function as a transcriptional repressor. Further *in vivo* experiments showed that phages dependent on *acrF1* and *acrF2* for viability[4] were markedly inhibited by overexpression of the Csy complex subunits (Fig. 2b). The elevated level of Csy proteins probably increases the number of active Csy complexes and/or binds and titrates out anti-CRISPR molecules, resulting in insufficient levels of anti-CRISPR proteins to support robust phage replication. Phages dependent on *acrF3* were not affected under these conditions because this anti-CRISPR protein binds to Cas3, the level of which is unchanged (overexpression of Cas3 inhibited cell growth). Interestingly, Csy subunit overexpression also inhibited a phage expressing *acrF4* (gene *37* from phage JBD26), an anti-CRISPR protein that could not be purified. In addition, expression of this anti-CRISPR in the transcriptional repression assay resulted in a blue–green culture (Fig. 2a). These complementary results imply that AcrF4 binds the Csy complex, which we have experimentally confirmed (Extended Data Fig. 5). We conclude that our *in vivo* experiments are able to distinguish the effects of anti-CRISPR proteins that inactivate the Csy complex from those that inhibit Cas3.

AcrF1 and AcrF2 both prevent DNA binding by the Csy complex, but might achieve this outcome through different mechanisms. The Csy complex assembles with a Csy1–Csy2 heterodimer bound at the 5′ end of the crRNA and a Csy4 monomer bound to the 3′ end, with six Csy3 subunits arrayed along the backbone of the spacer region in between (Fig. 3a)[6,8]. By purifying the Csy1–Csy2 heterodimer on its own and mixing it with purified anti-CRISPR proteins, we found that it co-eluted with AcrF2 in SEC experiments, but not with AcrF1 (Fig. 3b and

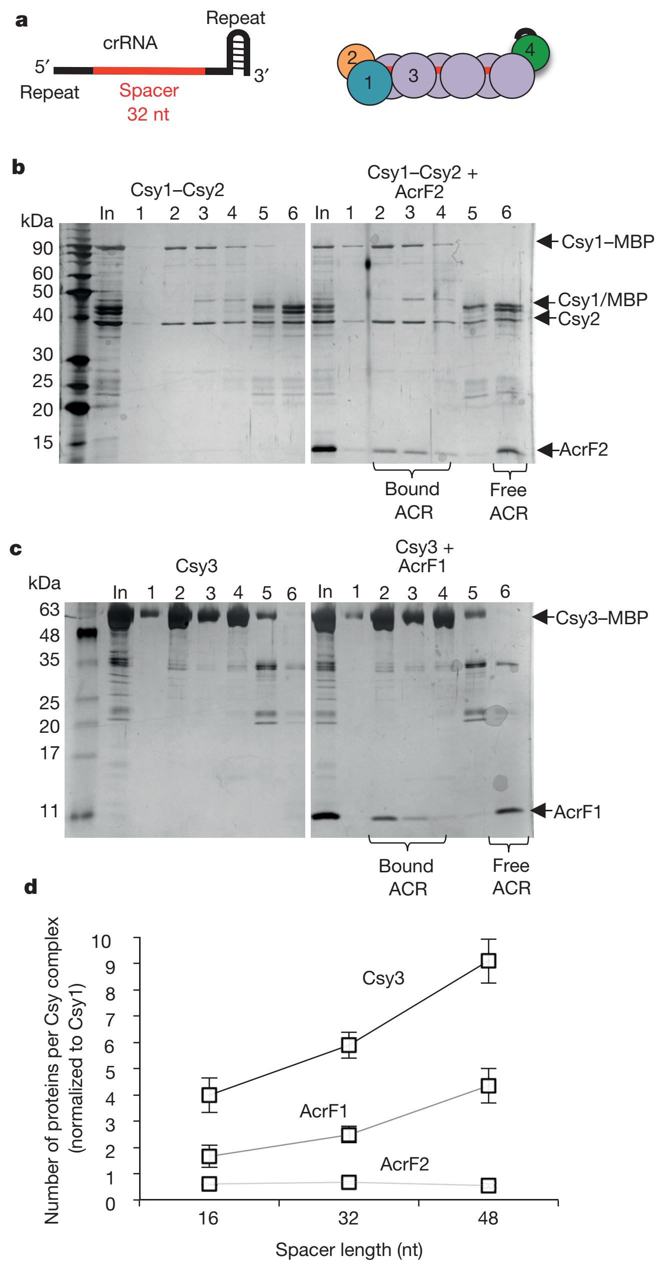


Figure 3 | AcrF1 and AcrF2 bind distinct Csy complex subunits. a, A schematic of the crRNA showing the repeat-derived regions of the crRNA (black) and the 32-nucleotide (nt) spacer region (red). The coloured circles represent the Csy1–4 subunits. **b**, **c**, Purified 6×His/MBP-tagged Csy1–Csy2 heterodimer (**b**) or Csy3 (**c**) was fractionated by SEC in the presence (right) or absence (left) of the indicated anti-CRISPR proteins. The SEC fractions were analysed by SDS–PAGE. The 'In' lanes show the protein mixture that was loaded onto the SEC column and fractions are numbered. **d**, Purified Csy complexes with 16-, 32-, or 48-nucleotide crRNA spacer regions were bound to AcrF1 or AcrF2 and fractionated by SEC. The stoichiometry of the bound anti-CRISPR proteins was quantified through densitometry of the Coomassie blue stained gels. An average of three independent experiments is shown with error bars representing s.d.

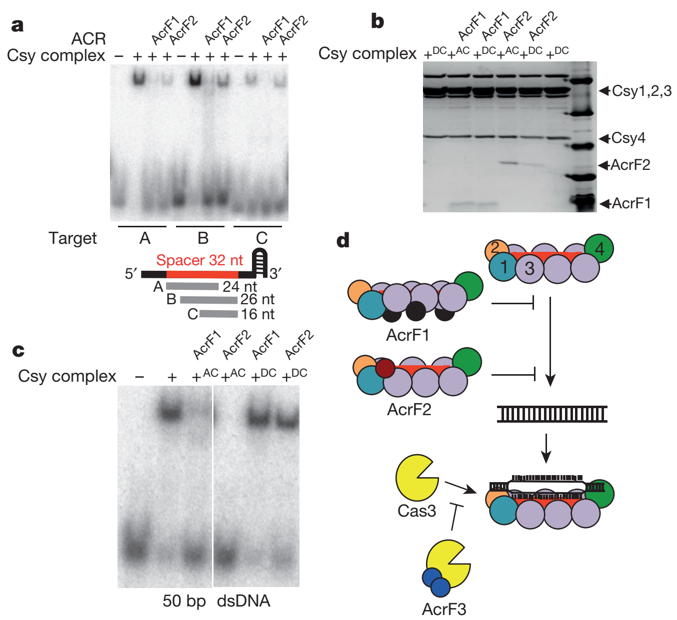


Figure 4 | Two anti-CRISPR proteins inhibit target recognition via unique mechanisms. a, EMSA experiments were used to assay binding of the Csy complex to three different ssDNA oligonucleotides (labelled A, B and C) that are complementary to different regions of the crRNA spacer as shown in the schematic (see Extended Data Fig. 9b). Where noted, the Csy complex was pre-incubated with the indicated anti-CRISPR. **b**, **c**, Apo–Csy complex (AC) or DNA-bound Csy complex (DC) was incubated with AcrF1 or AcrF2. **b**, This mixture was fractionated by SEC and fractions were visualized by SDS–PAGE. **c**, An EMSA experiment is shown with binding to dsDNA in the same experimental setup as in **b**. **d**, A model summarizing anti-CRISPR mechanisms. Arrows indicate the steps of the uninhibited CRISPR–Cas interference pathway. Numbers in the Csy complex indicate the Csy subunits. The lines with flat ends indicate the step in the CRISPR–Cas pathway blocked by each anti-CRISPR. The manner in which each anti-CRISPR binds to CRISPR–Cas components is also shown. AcrF1 makes the whole crRNA inaccessible while AcrF2 occludes the 5′ end.

Extended Data Fig. 6a). By contrast, AcrF1, but not AcrF2, bound Csy3 (Fig. 3c and Extended Data Fig. 6b). Csy3 eluted in monomeric and multimeric forms in SEC experiments, with AcrF1 binding predominantly to the multimeric fraction (Fig. 3c). The presence of distinct binding sites for AcrF1 and AcrF2 on the intact Csy complex was corroborated through competition experiments showing that both anti-CRISPR proteins could simultaneously bind the Csy complex and that the presence of one had no effect on the binding ability of the other (Extended Data Fig. 6c). RNase A treatment of the Csy complex, which resulted in Csy4 dissociation, had no effect on the binding of either anti-CRISPR (Extended Data Fig. 7). Quantification of the co-eluted fractions of AcrF1 or AcrF2 with the Csy complex by protein gel electrophoresis revealed the stoichiometry of AcrF1 to be 2.6 ± 0.3 proteins per Csy complex, while AcrF2 was 0.8 ± 0.1 (Extended Data Fig. 7c). To verify these stoichiometries, we created Csy complexes with shorter (16 nucleotides; Csy_{16} complex) and longer spacer regions (48 nucleotides; Csy_{48} complex). The purified Csy_{16} complex contained fewer molecules of Csy3 (4 ± 0.7) than wild type, and the Csy_{48} complex contained a proportionally greater number (9 ± 0.8) (Fig. 3d and Extended Data Fig. 8). Concomitant with the altered number of Csy3 molecules in the Csy_{16} and Csy_{48} complexes, we observed corresponding changes in the number of AcrF1 molecules bound, with the ratio of Csy3 to AcrF1 remaining constant. These results imply that AcrF1 binds along the full length of the Csy3 'spine' of the complex. Its binding sites are probably at the interaction interfaces of the Csy3 subunits, which would account for the 2:1 Csy3/AcrF1 stoichiometry and for AcrF1 binding to only the multimeric Csy3 fraction (Fig. 3c). In contrast to AcrF1, the number of AcrF2 molecules bound to the altered Csy complexes did not change as the number of Csy3 molecules increased or decreased, consistent with AcrF2 binding to the Csy1–Csy2 heterodimer.

To define further the sites of action of the anti-CRISPR proteins on the Csy complex, we performed DNA-binding assays using ssDNA molecules complementary to subregions of the crRNA spacer. As shown in Fig. 4a, AcrF1 inhibited binding to all the ssDNA molecules tested. By contrast, AcrF2 prevented binding to a 24-nucleotide ssDNA molecule complementary to the 5′ end of the crRNA, including the

seed region, but did not inhibit binding to a 16-nucleotide ssDNA complementary to the 3′ end of the spacer. Binding to a 26-nucleotide ssDNA binding the 3′ end was only partially inhibited. These data suggest that AcrF2 inhibits DNA binding by sterically blocking the 5′ end of the crRNA spacer through its interaction with Csy1–Csy2, which is expected to be bound to this region of the crRNA[17,18]. Addition of AcrF2 to a Csy complex that had been pre-saturated with target DNA resulted in an approximately 60% decrease in the binding level of this anti-CRISPR, suggesting that AcrF2 and DNA compete for an overlapping binding interface (Fig. 4b and Extended Data Fig. 9a). Consistent with this result, addition of AcrF2 to a DNA-bound Csy complex resulted in appreciably decreased DNA binding as detected by EMSA (Fig. 4c). Parallel experiments performed with AcrF1 showed that the binding of AcrF1 to the Csy complex was not affected by prior binding to DNA (Fig. 4b). We conclude that the interaction of AcrF1 with the full length of the spine of the complex formed by multiple Csy3 molecules and the crRNA accounts for its ability to block binding to all dsDNA and ssDNA molecules tested. Furthermore, the ability of AcrF1 and DNA to bind the Csy complex simultaneously suggests an allosteric mechanism for the activity of this anti-CRISPR. Thus, the mechanisms of AcrF1 and AcrF2 are distinct, using different Csy protein-binding partners, stoichiometry and DNA occlusion mechanisms (that is, steric versus allosteric).

We provide the first insight into the mechanisms by which proteins can inhibit a CRISPR–Cas system. The diverse and distinct mechanisms discovered here (Fig. 4d) reflect the deep evolutionary roots of the virus–host arms race. Anti-CRISPR proteins, both known[4,19] and yet to be discovered, will provide an extensive set of valuable tools both better to understand and to manipulate CRISPR–Cas systems. One example is our finding that AcrF3 converts the CRISPR–Cas system into a gene regulator by blocking Cas3 recruitment. Since CRISPR–Cas systems perform a variety of roles beyond destroying foreign DNA[20], many important functions may be fulfilled by proteins that interact with CRISPR–Cas components and thus alter the activity of the system.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 November 2014; accepted 29 July 2015.
Published online 23 September 2015.

1. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).
2. Makarova, K. S. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nature Rev. Microbiol.* **9,** 467–477 (2011).
3. Jore, M. M., Brouns, S. J. J. & van der Oost, J. RNA in defense: CRISPRs protect prokaryotes against mobile genetic elements. *Cold Spring Harb. Perspect. Biol.* **4,** a003657 (2012).
4. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493,** 429–432 (2013).
5. van der Oost, J., Westra, E. R., Jackson, R. N. & Wiedenheft, B. Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nature Rev. Microbiol.* **12,** 479–492 (2014).
6. Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J. A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329,** 1355–1358 (2010).
7. Wiedenheft, B. *et al.* RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl Acad. Sci. USA* **108,** 10092–10097 (2011).
8. van Duijn, E. *et al.* Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in clustered-regularly-interspaced shot-palindromic-repeats (CRISPR)-associated protein complexes from *Escherichia coli* and *Pseudomonas aeruginosa*. *Mol. Cell. Proteomics* **11,** 1430–1441 (2012).
9. Westra, E. R. *et al.* CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* **46,** 595–605 (2012).
10. Huo, Y. *et al.* Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nature Struct. Mol. Biol.* **21,** 771–777 (2014).
11. Rollins, M. F., Schuman, J. T., Paulus, K., Bukhari, H. S. T. & Wiedenheft, B. Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from *Pseudomonas aeruginosa*. *Nucleic Acids Res.* **43,** 2216–2222 (2015).
12. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl Acad. Sci. USA* **108,** 10098–10103 (2011).
13. Wurtzel, O. *et al.* The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. *PLoS Pathog.* **8,** e1002945 (2012).
14. Luo, M. L., Mullis, A. S., Leenay, R. T. & Beisel, C. L. Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res.* **43,** 674–681 (2015).
15. Rath, D., Amlinger, L., Hoekzema, M., Devulapally, P. R. & Lundgren, M. Efficient programmable gene silencing by Cascade. *Nucleic Acids Res.* **43,** 237–246 (2015).
16. Cady, K. C. & O'Toole, G. A. Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. *J. Bacteriol.* **193,** 3433–3445 (2011).
17. Jackson, R. N. *et al.* Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* **345,** 1473–1479 (2014).
18. Mulepati, S., Héroux, A. & Bailey, S. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* **345,** 1479–1484 (2014).
19. Pawluk, A., Bondy-Denomy, J., Cheung, V. H. W., Maxwell, K. L. & Davidson, A. R. A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. *MBio* **5,** e00896 (2014).
20. Westra, E. R., Buckling, A. & Fineran, P. C. CRISPR–Cas systems: beyond adaptive immunity. *Nature Rev. Microbiol.* **12,** 317–326 (2014).

Acknowledgements We thank W. Navarre and E. Westra for reading the manuscript. This work was supported by an Operating Grant to A.R.D. (MOP-130482) and to K.L.M. (MOP-136845), both of which were from the Canadian Institutes of Health Research (CIHR). J.B.-D. was supported by a CIHR Canada Graduate Scholarship Doctoral Award and an Ontario Graduate Scholarship award. Research in the Wiedenheft laboratory is supported by the National Institutes of Health (P20GM103500 and R01GM108888), the National Science Foundation EPSCoR (EPS-110134), the M.J. Murdock Charitable Trust, and the Montana State University Agricultural Experimental Station.

Author Contributions J.B.-D. designed, performed and supervised experiments and wrote the manuscript. B.G., S.S., M.D., and Y.H.-R. performed experiments. M.F.R. and B.W. provided essential reagents and experimental assistance. K.L.M. supervised experiments. A.R.D. designed and supervised experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.R.D. (alan.davidson@utoronto.ca) or K.L.M. (karen.maxwell@utoronto.ca).

METHODS

Protein purification. All proteins were affinity purified using Ni-NTA beads (Qiagen) to isolate recombinant proteins bearing a terminal 6×His tag. Anti-CRISPR proteins were expressed from the p15TV-L vector (NCBI accession number EF456736), which possesses a T7 promoter and an amino-terminal 6×His tag. Constructs expressing Csy1–4 containing a 6×His tag on either Csy3 or Csy4 were co-expressed with a construct producing a crRNA as previously described[7]. Individual Cas proteins (Csy1–Csy2, Csy3, and Cas3) were expressed from pHMGWA (NCBI accession number EU680841), which also has a T7 promoter. The proteins in this vector were tagged with a maltose-binding protein and 6×His.

Cultures of *E. coli* BL21 containing a plasmid expressing a protein of interest were grown to an optical density ($OD_{600\ nm}$) of 0.5 and then induced with 1 mM isopropyl-β-D-thiogalactoside (IPTG) for 3 h at 37 °C (anti-CRISPRs, Csy3) or for 16 h at room temperature (Csy complex, Csy1–Csy2, Cas3). Cells were collected by centrifugation at 5,000*g* for 10 min and resuspended in a binding buffer (20 mM Tris, pH 7.5, 250 mM NaCl, 5 mM imidazole, 1 mM dithiothreitol (DTT) and 1 mM PMSF). The cells were lysed by sonication and the resulting lysate was centrifuged at 15,000*g* for 15 min to remove cell debris. The supernatant was mixed with Ni-NTA beads that had been washed in binding buffer (without DTT) five times. Binding to the beads proceeded for 1 h at 4 °C under gentle rotation, at which point the lysate and beads were passed through a column, washed 3–5 times with binding buffer containing 30 mM imidazole and ultimately eluted in buffer containing 250 mM imidazole. Colourimetric Bradford assays were conducted during the procedure to determine the number of washes to perform and elution fractions to collect. Purified protein was dialysed into the binding buffer containing 5 mM imidazole to remove excess imidazole and visualized on Coomassie blue R250 stained SDS–PAGE gels. Cas3 was purified following the same general protocol but in a buffer optimized for this protein (50 mM HEPES, pH 7.5, 500 mM NaCl, 5% glycerol, 1 mM DTT, supplemented with 1 mM PMSF and 150 μM $NiSO_4$ in the lysis buffer). Purified Cas3 was concentrated and buffer exchanged in an Amicon Ultra centrifugal filter (Millipore) into a different buffer (20 mM HEPES, pH 7.5, 300 mM KCl, 5% glycerol, 1 mM DTT) for protein interaction assays. Csy1–Csy2 also purified in the same buffer as Cas3 (with $NiSO_4$ omitted). Purified Csy1–Csy2 was then dialysed into a different buffer (20 mM HEPES, pH 7.5, 250 mM NaCl, 5% glycerol, 1 mM DTT) for protein interaction experiments.

Size-exclusion chromatography. Affinity-purified proteins were fractionated by SEC using a GE Life Sciences Superdex 200 10/300 column. Fractions were collected in 0.5 ml volumes and monitored by optical density at 280 nm. SDS–PAGE gels were stained with silver nitrate or Coomassie blue R250 to identify proteins. In interaction experiments, purified proteins were mixed together before fractionation by SEC and co-eluting proteins were identified by SDS–PAGE. The Csy complex or Csy proteins and an anti-CRISPR protein of interest were generally incubated together for 1 h at 4 °C. This mixture was then applied to the SEC column at room temperature. A fraction of the input (~0.5%) was also kept for SDS–PAGE analysis.

Anti-CRISPR stoichiometry. The purified Csy complex was incubated with ~10-fold molar excess of purified anti-CRISPR proteins. This mixture was fractionated by SEC as described earlier. The Csy complex peak fraction was run on SDS–PAGE gels in twofold serial dilutions. The protein bands were identified with Coomassie blue R250. Image Lab Software (Bio-Rad) was used to quantify band intensities and calculate the relative stoichiometries of the various subunits and anti-CRISPRs, after adjusting for molecular weight and comparing dilutions. Our estimates of the absolute stoichiometries of the Csy subunits is based on the stoichiometry of the Csy complex established in previous publications[7,8].

RNase A treatment of the Csy complex. Pancreatic RNase A (73 μM) was used to treat the Csy complex (4 μM) for 30 min at 37 °C. After digestion, the treated Csy complex was fractionated by SEC in the absence or presence of an anti-CRISPR protein. Fractions from SEC were analysed on Coomassie stained SDS–PAGE gels to visualize proteins and SYBR Gold stained TBE-Urea gels to visualize nucleic acid.

Isothermal titration calorimetry. Purified Csy complex was added to the isothermal titration calorimetry (ITC) chamber at a concentration of 7.5 μM. The DNA ligand (8-nucleotide ssDNA) was placed in the injection syringe at a concentration of 75 μM. After a null injection of 0.3 μl of titrant, 3 μl of titrant were injected 13 times, with 120 s intervals between the injections to establish a baseline. The DNA titrant and Csy complex were in the same buffer (20 mM Tris, pH 7.5, 250 mM NaCl, 5 mM imidazole) and the experiment was temperature controlled at 25 °C. To assess the role of AcrF1 in interfering with the interaction between the Csy complex and a DNA target, the Csy complex was first incubated with a ~10-fold molar excess of anti-CRISPR proteins for 1 h at 4 °C. This mixture was then applied to the chamber, the temperature equilibrated to 25 °C and the DNA titration performed.

Electrophoretic mobility shift assay. A 50-nucleotide ssDNA molecule was synthesized (Eurofins Genomics) that contains 32 nucleotides of complementarity to the crRNA in the purified Csy complex. The DNA (200 nM) was phosphorylated in a T4 polynucleotide kinase reaction with [γ-^{32}P]ATP. The reaction was stopped with 12 mM EDTA and GE MicroSpin G-25 columns were used to remove remaining radiolabelled nucleotides. To generate dsDNA, the labelled strand was heated to 98 °C in the presence of a twofold excess of an unlabelled complementary strand and allowed to return slowly to room temperature. Csy complex–DNA-binding reactions were conducted in a binding buffer (10 mM HEPES, pH 7.5, 1 mM $MgCl_2$, 20 mM KCl, 1 mM TCEP, bromophenol blue and 6% glycerol) at 37 °C for 15 min. The concentration of the Csy complex used in EMSA experiments varied, depending on the oligonucleotide target being used. For 50 bp dsDNA EMSA reactions, 100 nM of the Csy complex was routinely used in reactions, with <1 nM labelled DNA. Anti-CRISPR proteins were used at a tenfold molar excess compared to the Csy complex and allowed to incubate with Apo–Csy complex or DNA-bound Csy complex for 1 h. After the appropriate incubation, the reactions were resolved on native 6% polyacrylamide TBE gels. Gels were wrapped in Saran wrap and visualized with a phosphoscreen and Typhoon imager. Optimal exposures were ~2–3 h.

For EMSA experiments involving Cas3, the Csy complex and target DNA were prebound as described above. 6×His-tagged Cas3 was purified by Ni-NTA chromatography (6×His) followed by SEC, concentrated, transferred into the EMSA reaction buffer, flash frozen in small volumes (50 μl) and stored at −70 °C. Cas3 was added to the EMSA reaction at a final concentration of 400 nM and incubated for 30 min at 37 °C. ATP was added at a final concentration of 2 mM and all reactions with Cas3 also contained 100 μM $CoCl_2$.

Pyocyanin repression. A crRNA was designed to target the promoter region of *phzM*, a gene required for the biosynthesis of the blue–green pigment pyocyanin. Two complementary oligonucleotides were synthesized containing two 28 bp PA14 CRISPR repeat sequences, flanking a 32 bp sequence with perfect complementarity to the −35/−10 region of the *phzM* promoter (position 813576–813607 in the PA14 genome). The spacer was designed to produce a crRNA that would bind to the non-template strand, in a position where the protospacer adjacent motif (GG) is present. The oligonucleotides were annealed and cloned into an arabinose inducible *P. aeruginosa* expression vector, pHERD30T. This construct was then used to transform PA14 strains possessing single *cas* gene knockouts or wild-type PA14 possessing prophages expressing various anti-CRISPRs. Individual transformants were grown overnight (~20 h) in 2 ml of King's A media in 50 μg ml^{-1} gentamicin and 0.025% arabinose, to induce expression of the crRNA. Pyocyanin was extracted with an equal volume of chloroform, and then mixed with 1 ml of 0.2 M HCl, producing a pink–red colour proportional to the amount of pyocyanin, which was quantitated by measuring absorbance at 520 nm. Anti-CRISPR proteins were expressed from the following prophages: JBD30 (AcrF1), D3112 (AcrF2), JBD26 (AcrF4), JBD5 (AcrF3 and AcrF5), and JBD88a (AcrF3H). Since phage JBD5 contains two type I-F anti-CRISPR proteins, phage JBD88a (possessing a homologue of AcrF3 with 86% protein sequence identity) was also used.

Competition experiments. To determine whether the two anti-CRISPR proteins that bind to the Csy complex compete with each other for the same binding site, the first anti-CRISPR was added for 1 h at 4 °C and then the second for the same amount of time. This entire mixture was then fractionated by SEC.

To determine whether DNA and anti-CRISPR proteins compete for the same binding site, the purified Csy complex (4.5 μM) was mixed with a 50 bp dsDNA target (10 μM) and incubated for 15 min at 37 °C in the same buffer in which the proteins were purified (20 mM Tris, pH 7.5, 250 mM NaCl, 5 mM imidazole). This DNA-bound Csy complex was then mixed with a tenfold molar excess of AcrF1, AcrF2, or an equivalent volume of buffer and incubated for 1 h at 4 °C. This mixture was fractionated by SEC. The fraction containing the Csy complex was analysed on Coomassie blue stained SDS–PAGE gels or SYBR Gold stained TBE-Urea gels.

Plaque assays with Csy subunit overexpression. To assess the consequence of Csy protein overexpression on phages possessing distinct anti-CRISPR proteins *in vivo*, a pHERD30T derived plasmid expressing the *csy1*, *csy2*, *csy3* and *csy4* genes was used to transform *P. aeruginosa* strain PA14. Phage lysates were spotted in tenfold serial dilutions onto a lawn of PA14 containing empty vector, or the plasmid expressing the *csy* genes. Phages JBD30, JBD26, D3112 and JBD88a all have protospacers that display 100% identity to spacers 17 and 20 in the PA14 *CRISPR2* locus[4]. JBD5 has a protospacer matching *CRISPR2* spacer 1 that has been shown to be targeted[4,21].

RT–qPCR. RT–qPCR reactions were conducted as described previously[4]. Briefly, total RNA was extracted and DNase treated. One nanogram of total RNA was subjected to a reverse transcription reaction and qPCR, using primers specific to *phzM* or a control, *rpsL*. The efficient removal of DNA from the RNA

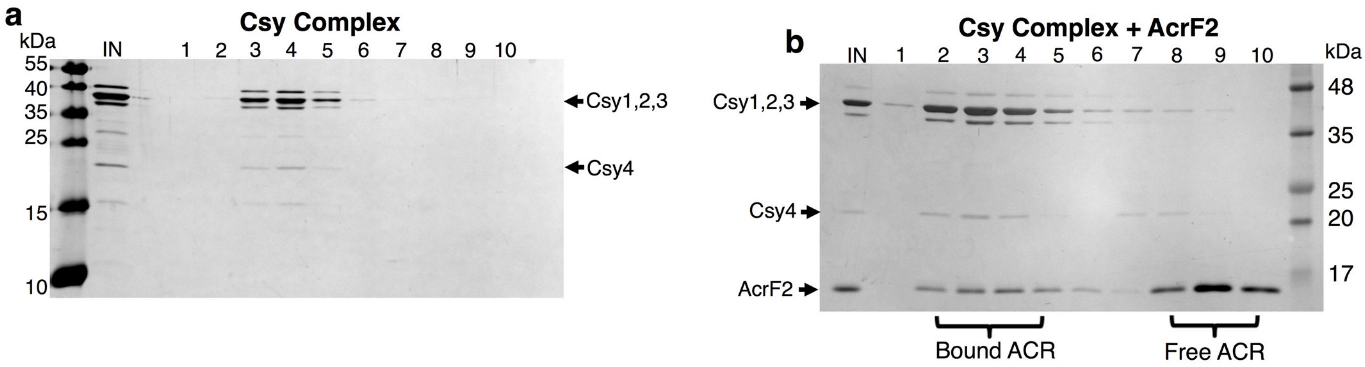
preparation was confirmed by including controls for each sample without reverse transcriptase added.

AcrF2 misannotation. The D3112 phage genome has an annotated open reading frame identified as gene *30*, which is a predicted 90 amino acid protein (NCBI accession number NC_005178). This version of the gene was previously identified as an anti-CRISPR, although overexpression from a plasmid was required for activity[4]. A nucleotide alignment of the anti-CRISPR region of many phages revealed that all phage anti-CRISPR operons possess a start codon (ATG) at the same position for the first anti-CRISPR gene, except phage D3112. Phages D3112 and MP29 (which has a D3112 gene *30* homologue), had the start position annotated downstream of this commonly used ATG, at a second ATG, in frame with the first, resulting in a putative truncation of six amino acid residues. Re-cloning of the gene to include these six residues resulted in a construct that had full anti-CRISPR activity in the absence of overexpression. Thus, this 96-residue protein (sequence shown later, with new residues in bold) is the version that was used in all downstream experiments presented here and in affinity purification, after addition of the appropriate tag. All other anti-CRISPR protein sequences are as reported in ref. 4. AcrF2: **MTKTAQ**MIAQQHKDTVAACEAAEAIAIAKDQVWDGEGYT KYTFDDNSVLIQSGTTQYAMDADDADSIKGYADWLDDEARSAEASEIER LLESVEEE.

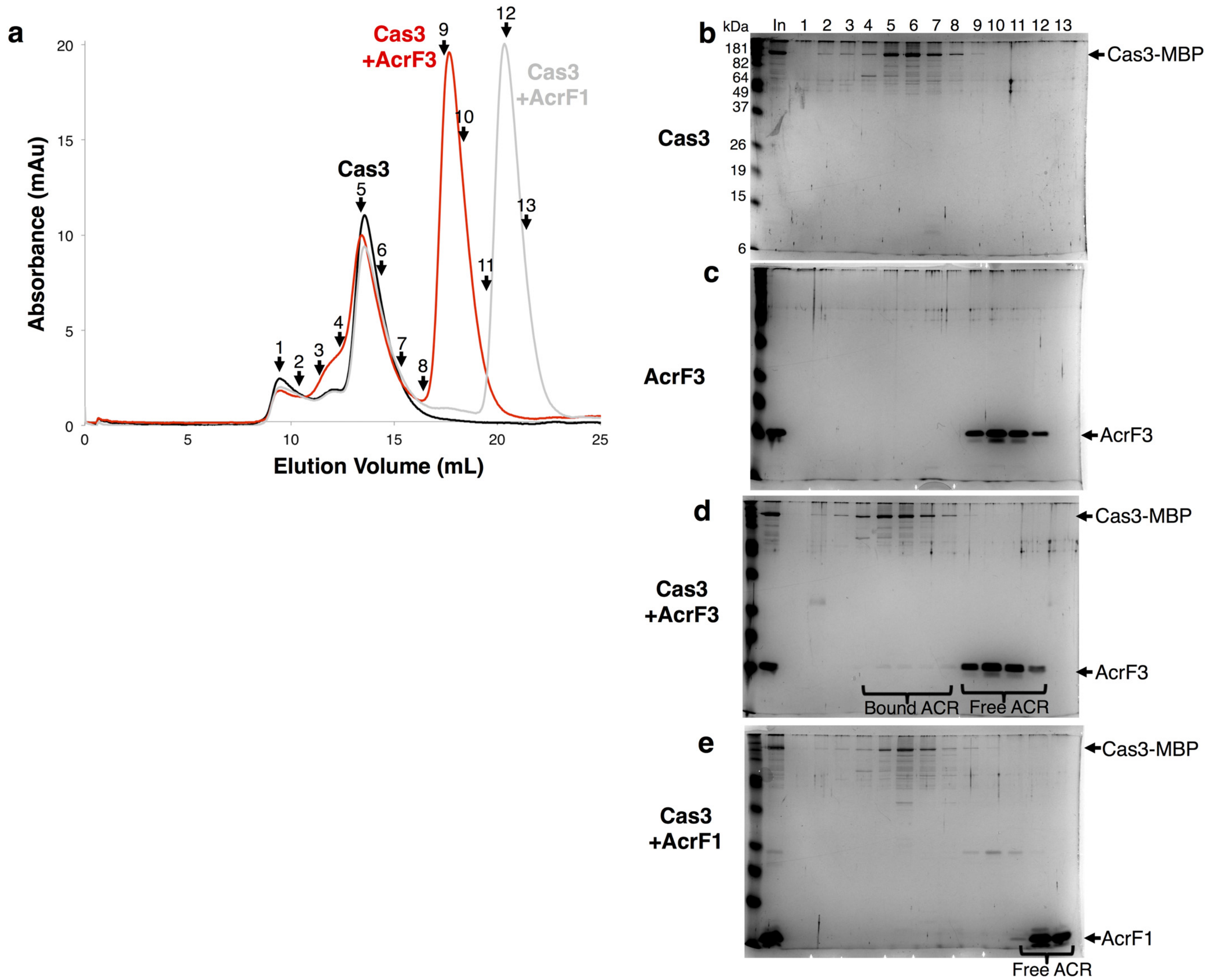
Statistics, reagents and data deposition. To assess interactions between anti-CRISPR proteins and the Csy complex or purified Cas proteins, mixed components were fractionated by SEC. Each result shown in the manuscript was obtained on at least two independent occasions. ITC, EMSA and plaque assays were all replicated at least three times. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

The sequences of the anti-CRISPR proteins are present in ref. 4, with full genomes for phages JBD30, D3112, JBD5, JBD26 and JBD88a available on NCBI (accession numbers: NC_020198, NC_005178, NC_020202, JN811560 and NC_020200, respectively).

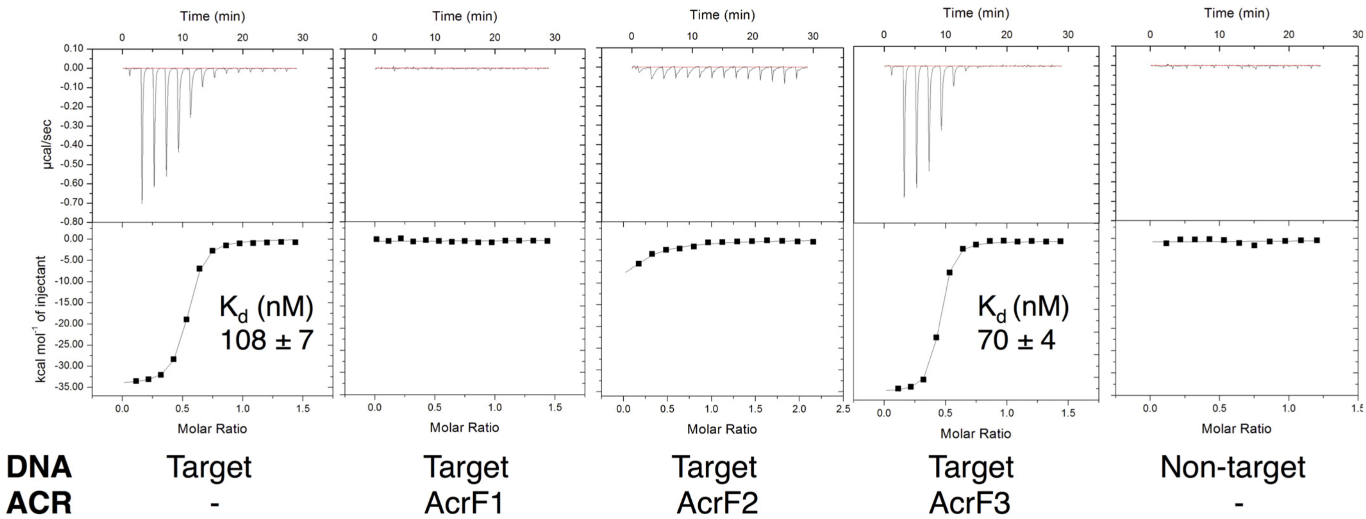
21. Cady, K. C., Bondy-Denomy, J., Heussler, G. E., Davidson, A. R. & O'Toole, G. A. The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. *J. Bacteriol.* **194,** 5728–5738 (2012).



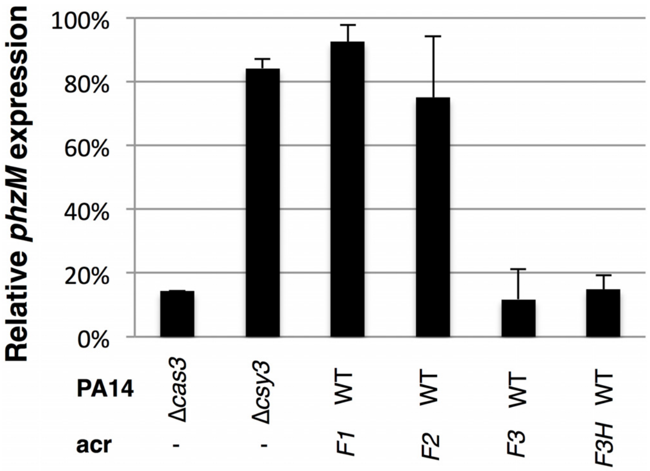
Extended Data Figure 1 | AcrF2 interacts with the Csy complex. **a**, **b**, Purified Csy complex was fractionated by SEC alone (**a**) or in the presence of AcrF2 (**b**). Fractions were analysed on a silver nitrate stained SDS–PAGE gel. The input (IN) and fractions are shown.



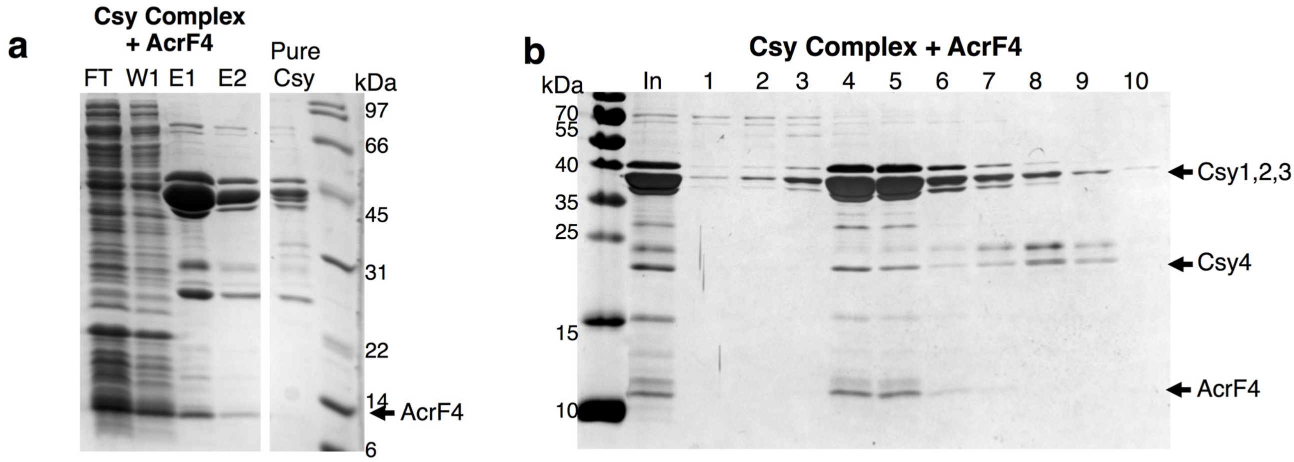
Extended Data Figure 2 | AcrF3, not AcrF1, interacts with Cas3. **a**, Cas3 was fractionated by SEC alone or in the presence of AcrF3 or AcrF1. Overlays of plots of elution volume versus optical density at 280 nm of the column eluates are shown. The numbers represent the fractions that were selected for analysis. **b**–**e**, Silver nitrate stained SDS–PAGE gels are shown from SEC experiments with Cas3 (**b**), AcrF3 (**c**), Cas3 with AcrF3 (**d**) or Cas3 with AcrF1 (**e**). The sample that was loaded onto the SEC column is shown as input (In) and fractions from the same elution positions are indicated numerically. AcrF3 is seen eluting in fractions 4–8 only in the presence of Cas3. There is also a visible shift in the Cas3 elution profile in the presence of AcrF3 but not AcrF1 (fractions 3–5).



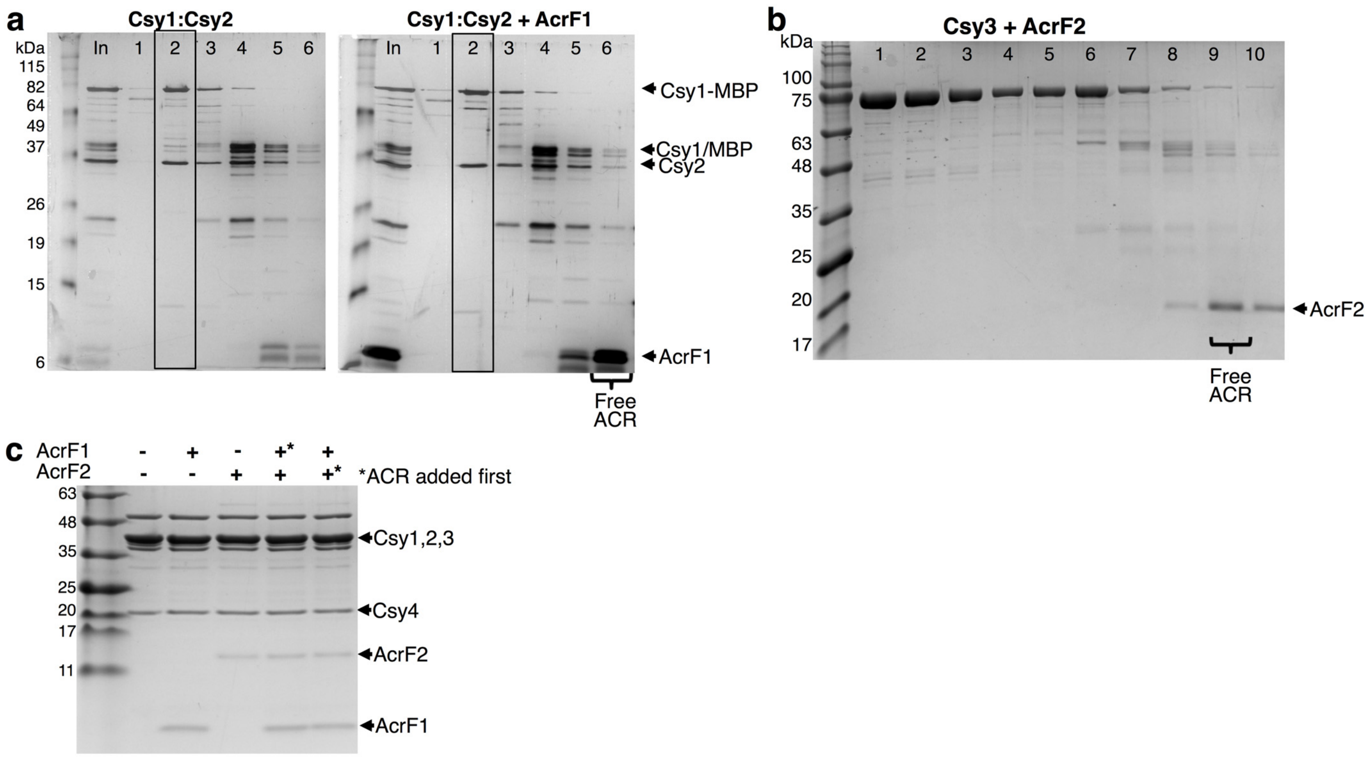
Extended Data Figure 3 | AcrF1 and AcrF2 prevent target recognition by the Csy complex. Isothermal titration calorimetry (ITC) assays showing the Csy complex binding to an 8-nucleotide ssDNA target that comprises the seed region. No binding is observed in the presence of AcrF1, AcrF2 or with a non-target (the reverse complement sequence of the target) ssDNA substrate. A representative run is shown for each condition with the dissociation constant (K_d) value and error of fit from that particular run. Over multiple runs ($n = 6$) with the Csy complex binding to the ssDNA ligand, the average K_d value was 90 nM ± 37.



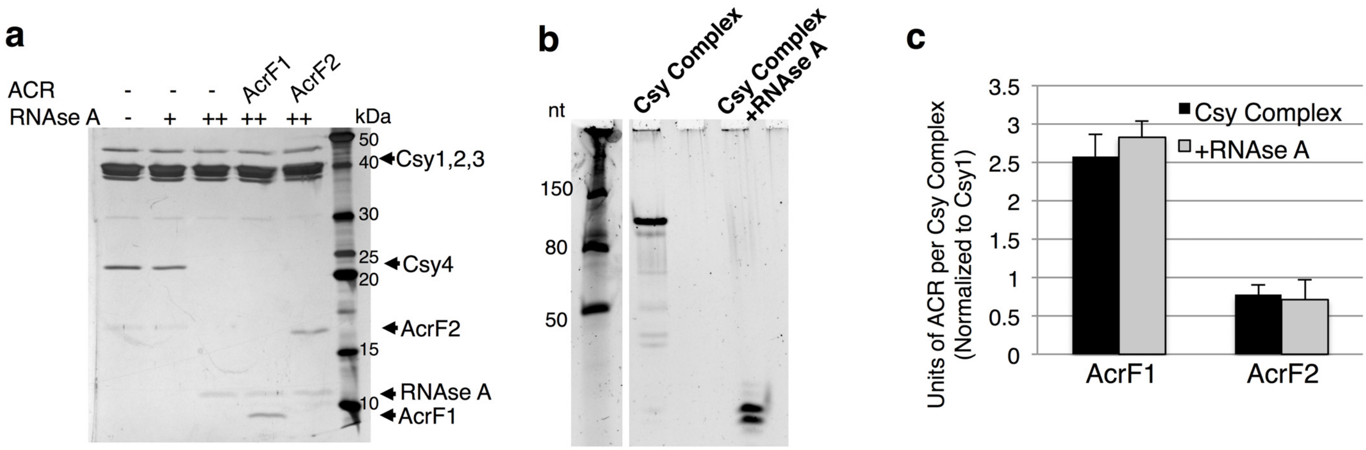
Extended Data Figure 4 | Expression of *phzM* is repressed by the Csy complex. The Csy complex was targeted to the promoter of the gene *phzM*, and repression efficiency was assayed by RT–qPCR (see Methods). The per cent repression of *phzM* in the indicated strains expressing a *phzM*-targeting crRNA relative to wild-type (WT) PA14 with an empty plasmid is shown. All values were normalized to *rpsL*, a gene encoding a ribosomal protein. Means ± s.d. are shown.



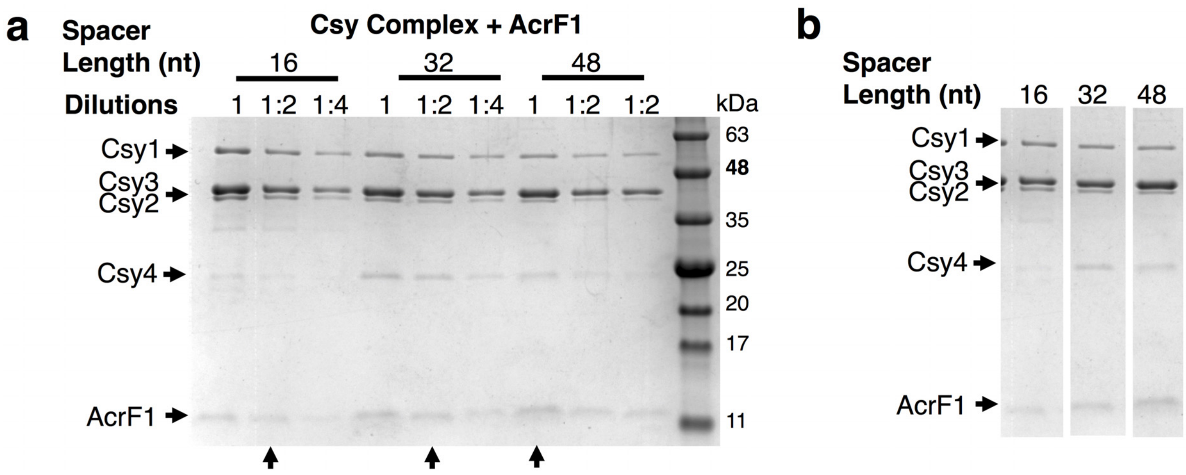
Extended Data Figure 5 | AcrF4 interacts with the Csy complex. Untagged AcrF4 was expressed in *E. coli* BL21 cells and a crude lysate of these cells was mixed with the Csy complex bound to Ni-NTA beads via a 6×His tag on Csy3. **a**, The flow through (FT), wash 1 (W1), and two elution fractions (E1, E2) from the Ni-NTA column are shown, as well as a comparison to pure Csy complex. **b**, The Ni-NTA elution fractions were fractionated by SEC, demonstrating a stable interaction between the Csy complex and AcrF4. The input (In) lane shows the sample that was loaded on the SEC column and numbered fractions are analysed on SDS–PAGE gels.



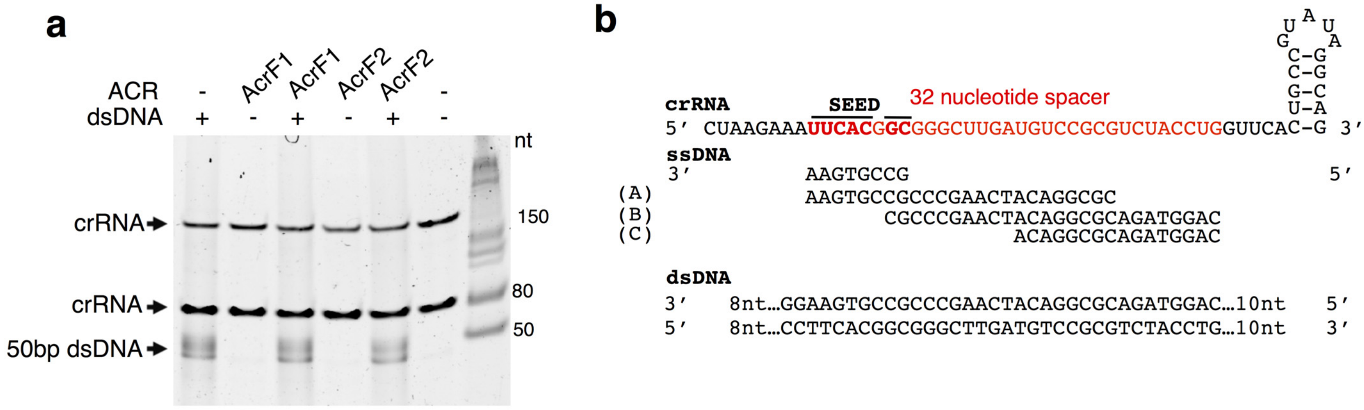
Extended Data Figure 6 | AcrF1 and AcrF2 bind the Csy complex at distinct locations. **a**, Purified Csy1–Csy2 heterodimer with an MBP and 6×His tag fused to Csy1 was fractionated by SEC in the presence or absence of AcrF1 (boxes indicate the Csy1–Csy2 peak). **b**, Purified MBP/6×His-tagged Csy3 was fractionated in the presence or absence of AcrF2. These are complementary experiments to those seen in Fig. 3b and c, respectively. Input (In) and selected fractions are shown on SDS–PAGE gels. **c**, AcrF1 and AcrF2 were incubated with the Csy complex singly or in combination. Asterisks designate which anti-CRISPR was added first to the reactions containing both anti-CRISPR proteins. The addition order did not affect the result since there is no competition for binding sites between these two anti-CRISPR proteins. After incubation, each mixture was fractionated by SEC and the peak Csy complex fraction is shown on an SDS–PAGE gel. In each experiment the anti-CRISPR proteins are in excess relative to the Csy complex.



Extended Data Figure 7 | AcrF1 and AcrF2 interact with an RNase-A-treated Csy complex. **a**, The Csy complex was treated with a low concentration (600 nM, +) of RNase A or a high concentration of RNase A (70 μM, ++). This mixture was fractionated by SEC, revealing Csy4 dissociation at the higher RNase A concentration. Pre-treatment of the Csy complex with RNase A, with the subsequent addition of AcrF1 or AcrF2 followed by SEC fractionation was then conducted. Peak Csy complex fractions are shown on an SDS–PAGE gel. **b**, A TBE-urea denaturing gel is shown, stained with SYBR gold, showing the native crRNA in the Csy complex and the protected fragments remaining after 70 μM RNase A treatment. **c**, Quantification of Coomassie blue stained gels from three independent preparations of the respective proteins is shown. Anti-CRISPR proteins bound with unaltered stoichiometry to RNase-A-pre-treated Csy complexes. Error bars represent s.d.



Extended Data Figure 8 | Twofold dilutions used to quantify anti-CRISPR binding stoichiometry. Csy complexes with crRNA molecules possessing spacers of differing lengths (16, 32, or 48 nucleotides) were purified and fractionated by SEC in the presence of AcrF1. A representative Coomassie blue stained SDS–PAGE gel is shown, with twofold dilutions of the peak fraction containing the Csy complex and co-eluting AcrF1. Arrows on the bottom of the gel indicate comparable dilutions based on the levels of Csy1. Note the increasing abundance of Csy3 and AcrF1. **b**, Lanes with arrows from the gel in **a** are shown next to each other for comparison.



Extended Data Figure 9 | dsDNA binds to the Csy complex after SEC fractionation. **a**, The same samples from Fig. 4a were run on a denaturing TBE-urea gel, stained with SYBR gold, to reveal the crRNA (two species are apparent), and the Csy-complex-bound 50 bp dsDNA. In these experiments, DNA was prebound to the Csy complex, and AcrF1 or AcrF2 were subsequently added to the DNA-saturated Csy complex. This mixture was then fractionated by SEC and the Csy-complex-containing peak fractions were analysed. **b**, A schematic showing the crRNA sequence with repeat-derived regions shown in black and the variable 32-nucleotide spacer region in red. The seed-interacting region that is critical for target recognition (nucleotides 1–5, 7, 8) is in bold. DNA oligonucleotides used in this study are shown, with labels 'A', 'B' and 'C' corresponding to the targets shown in Fig. 4c. The 8-nucleotide ssDNA substrate was used in ITC experiments (Extended Data Fig. 3), and the 50 bp dsDNA in EMSAs (Figs 1d and 4b).

DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins

Je Wook Woo[1,7], Jungeun Kim[2,3,7], Soon Il Kwon[1], Claudia Corvalán[4], Seung Woo Cho[3,6], Hyeran Kim[2], Sang-Gyu Kim[2], Sang-Tae Kim[2], Sunghwa Choe[1,4,5] & Jin-Soo Kim[2,3]

Editing plant genomes without introducing foreign DNA into cells may alleviate regulatory concerns related to genetically modified plants. We transfected preassembled complexes of purified Cas9 protein and guide RNA into plant protoplasts of *Arabidopsis thaliana*, tobacco, lettuce and rice and achieved targeted mutagenesis in regenerated plants at frequencies of up to 46%. The targeted sites contained germline-transmissible small insertions or deletions that are indistinguishable from naturally occurring genetic variation.

Programmable nucleases, such as zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and RNA-guided endonucleases (RGENs), have been used for genome editing in multiple cells and species including plants[1–3], paving the way for novel applications in biomedical research, medicine and biotechnology[4]. CRISPR RGENs are rapidly superseding ZFNs and TALENs owing to their ease of use; RGENs that consist of the Cas9 protein derived from *Streptococcus pyogenes* and guide RNAs (gRNAs) can be customized by replacing only the RNA component, sidestepping the labor-intensive and time-consuming protein engineering needed to customize TALENs and ZFNs. Programmable nucleases, delivered into plant cells either by using *Agrobacterium tumefaciens* or by transfecting plasmids that encode them, cleave chromosomal target sites in a sequence-dependent manner, producing site-specific DNA double-strand breaks (DSBs). The repair of these DSBs by endogenous systems results in targeted genome modifications.

It remains unclear whether genome-edited plants will be regulated by genetically modified organism (GMO) legislation in the EU and other regions[5]. Programmable nucleases induce small insertions and deletions (indels) or substitutions at chromosomal target sites that are indistinguishable from naturally occurring genetic variation. However, mutated plants might be considered GMOs by regulatory authorities in certain countries, which could reduce the potentially widespread use of programmable nucleases in plant biotechnology and agriculture. For example, when *A. tumefaciens* is used as a delivery vector, the resulting genome-edited plants contain foreign DNA sequences, including those that encode the programmable nucleases, in the host genome. Removal of these *A. tumefaciens*–derived DNA sequences by breeding is not feasible in species such as grape, potato and banana that reproduce asexually.

Non-integrating plasmids could be transfected into plant cells to deliver programmable nucleases. However, transfected plasmids are degraded in cells by endogenous nucleases, and the resulting small DNA fragments are sometimes inserted at both on-target and off-target sites in host cells[6]; therefore, this approach might be unsuitable in plants if regulatory approval is required.

Delivery of preassembled Cas9 protein-gRNA ribonucleoproteins (RNPs), rather than plasmids that encode these components, into plant cells could remove the likelihood of inserting recombinant DNA in the host genome[7]. Furthermore, as has been shown in cultured human cells[6], RGEN RNPs cleave chromosomal target sites immediately after transfection and are degraded rapidly by endogenous proteases in cells, which might reduce the frequency of mosaicism and off-target effects in regenerated whole plants. Because there is no need to optimize codon usage or find promoters that will express Cas9 and gRNAs when using protein-and-RNA-only systems, the use of preassembled RGEN RNPs could broaden the applicability of genome editing to all transformable plant species. In addition, using RGEN RNPs enables *in vitro* prescreening to guide the choice of highly active gRNAs[8] and genotyping of mutant clones via restriction fragment length polymorphism (RFLP) analysis[9]. To the best of our knowledge, RGEN RNPs have not been used in any plant species.

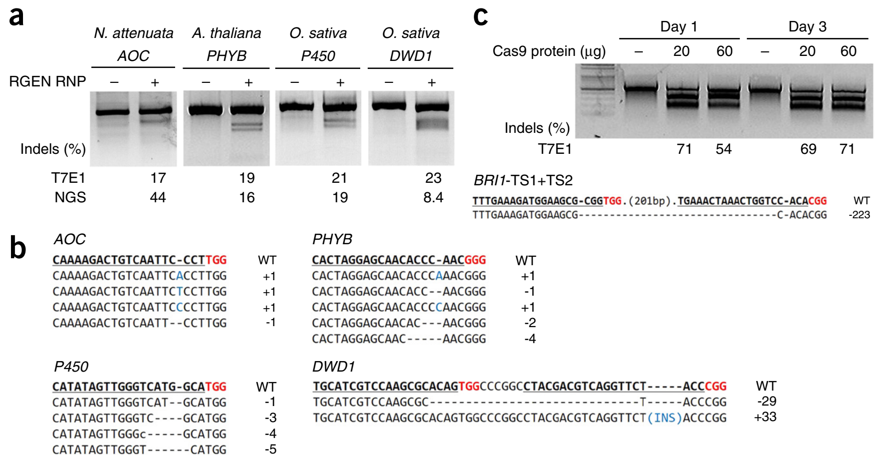
Here we report the delivery of RGEN RNPs into protoplasts of various plant species and the induction of targeted genome modifications in whole plants regenerated from them. Purified Cas9 protein was mixed with a two- to tenfold molar excess of gRNAs targeting four genes from three plant species *in vitro* to form preassembled RNPs. The RGEN RNPs were incubated with protoplasts derived from *A. thaliana*, tobacco (*Nicotiana attenuata*) and rice (*Oryza sativa*) in the presence of polyethylene glycol (PEG). We used both the T7 endonuclease I (T7E1)(ref. 10) assay and targeted deep sequencing to measure mutation frequencies in transfected cells (**Fig. 1a**,**b**). Indels were detected at the expected positions, that is, 3 nucleotides (nt) upstream of an NGG protospacer–adjacent motif (PAM), with frequencies that ranged from 8.4% to 44% (**Fig. 1a**).

We also co-transfected two gRNAs whose target sites were separated by 201 base pairs (bp) in the *BRASSINOSTEROID INSENSITIVE 1* (*BRI1*) gene in *Arabidopsis* to investigate whether the repair of two concurrent DSBs would result in targeted deletion of the intervening sequence, as has been seen in human cells[11]. Sanger sequencing showed that a 223-bp DNA sequence was deleted in protoplasts (**Fig. 1c**).

[1]Convergence Research Center for Functional Plant Products, Advanced Institutes of Convergence Technology, Yeongtong-gu, Suwon-si, Gyeonggi-do, Korea. [2]Center for Genome Engineering, Institute for Basic Science, Seoul, South Korea. [3]Department of Chemistry, Seoul National University, Seoul, South Korea. [4]School of Biological Sciences, College of Natural Sciences, Seoul National University, Seoul, South Korea. [5]Plant Genomics and Breeding Institute, Seoul National University, Seoul, South Korea. [6]Present address: Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California, USA. [7]These authors contributed equally to this work. Correspondence should be addressed to S.C. (shchoe@snu.ac.kr) or J.-S.K. (jskim01@snu.ac.kr).

Received 19 May; accepted 23 September; published online 19 October 2015; doi:10.1038/nbt.3389

Figure 1 RGEN RNP-mediated gene disruption in plant protoplasts of *Nicotiana attenuata, Arabidopsis thaliana* and *Oryza sativa.* (**a**) Mutation frequencies measured by the T7E1 assay and targeted deep sequencing. "–" indicates controls without RGEN RNP. (**b**) Mutant DNA sequences induced by RGEN RNPs in plant cells. The PAM sequences are shown in red and inserted nucleotides in blue. WT, wild-type. (**c**) A time-course analysis of genome editing of the *BRI1* gene in *A. thaliana* protoplasts. Top, the T7E1 assay ("–" indicates controls without Cas9 protein). Bottom, DNA sequences of the wild-type (WT) and mutant sequences.



Notably, RGEN-induced mutations were detected 24 h after transfection, suggesting that RGENs cut target sites immediately after transfection and induce mutation before a full cycle of cell division is completed.

Next, we investigated whether RGEN RNPs induce off-target mutations at sites highly homologous to on-target sites. We searched for potential off-target sites of the *PHYTOCHROME B* (*PHYB*) and *BRI1* gene–specific sgRNAs in the *Arabidopsis* genome using the Cas-OFFinder program[12] and used targeted deep sequencing to measure mutation frequencies (**Supplementary Fig. 1**). Indels were not detected at any sites that differed from on-target sites by 2–5 nt, in line with previous findings in human cells[13,14].

Finally, we transfected an RGEN RNP to disrupt the lettuce (*Lactuca sativa*) homolog of the *A. thaliana BRASSINOSTEROID INSENSITIVE 2* (*BIN2*) gene (**Supplementary Fig. 2**), which encodes a negative regulator in a brassinosteroid (BR) signaling pathway[15], into lettuce protoplasts, and obtained microcalli regenerated from the RNP-transfected cells (**Fig. 2** and **Supplementary Fig. 3**). We used the same RGEN RNP in an RFLP analysis to genotype the lettuce microcalli. Unlike the T7E1 assay, this analysis distinguishes monoallelic mutant clones (50% cleavage) from heterozygous biallelic mutant clones (no cleavage) and homozygous biallelic mutant clones (no cleavage) from wild-type clones (100% cleavage)[9]. Furthermore, the RGEN-RFLP assay is not limited by sequence polymorphisms near the nuclease target site that may exist in the lettuce genome. This assay showed that 2 of 35 (5.7%) calli contained monoallelic mutations and 14 of 35 (40%) calli contained biallelic mutations at the target site (**Fig. 2b**), demonstrating that RGEN-induced mutations were maintained after regeneration. The overall mutation frequency in lettuce calli was 46%. We used targeted deep sequencing to confirm genotypes in the 16 mutant calli. The number of base pairs deleted or inserted at the target

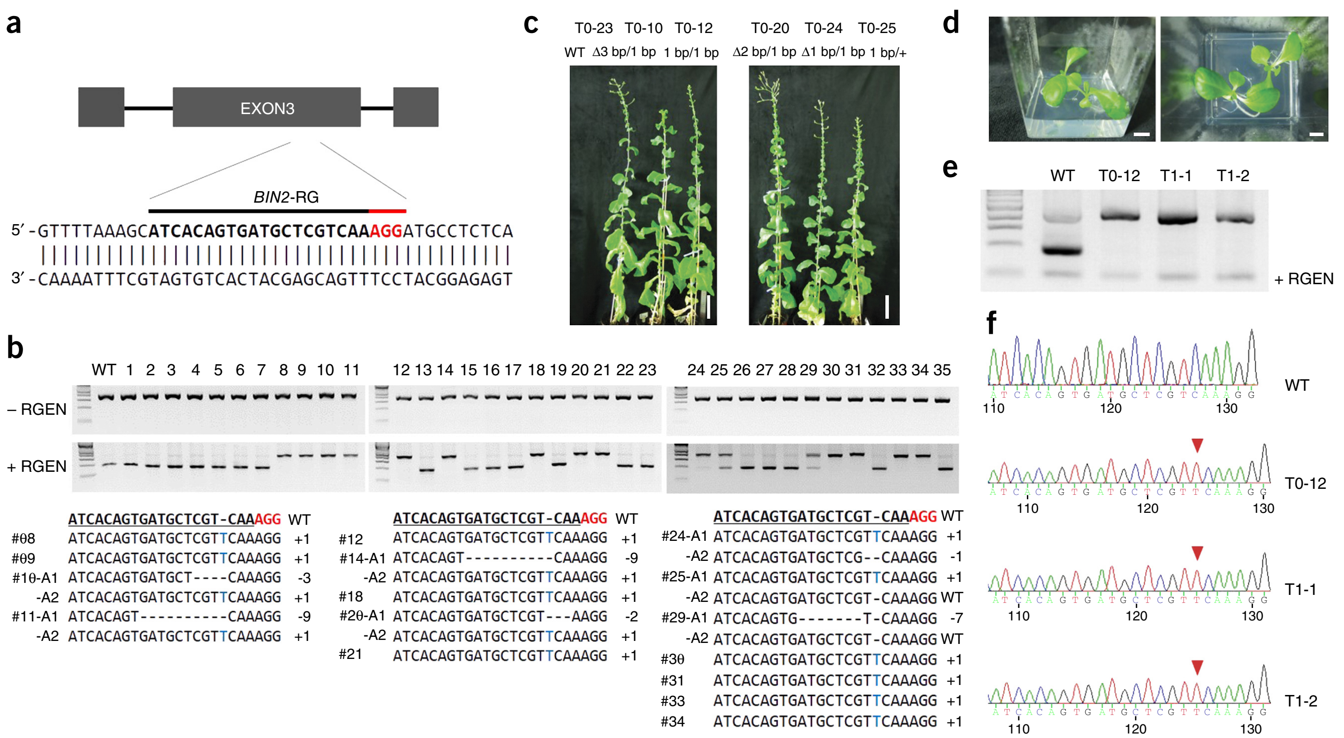


Figure 2 Targeted gene knockout in lettuce using an RGEN RNP. (**a**) The target sequence in the *BIN2* gene. The PAM sequence is shown in red. (**b**) Genotyping of microcalli. Top, RGEN-RFLP analysis. Bottom, mutant DNA sequences in microcalli. (**c**) Whole plants regenerated from RGEN RNP-transfected protoplasts. Scale bars, 10 cm. (**d**) T1 plantlets obtained from a homozygous biallelic mutant termed T0-12. Scale bars, 1 cm. (**e**) RGEN-RFLP analysis for genotyping T1 plantlets. (**f**) DNA sequences of the wild type, the T0-12 mutant, and T1 mutants derived from the T0-12 line. Red triangles indicate an inserted nucleotide.

site ranged from −9 to +1, consistent with mutagenic patterns observed in human cells[14]. No apparent mosaicism was detected in these clones (**Supplementary Fig. 4**), suggesting that the RGEN RNP cleaved the target site immediately after transfection and induced indels before cell division was completed.

We next evaluated whether the *BIN2*-specific RGEN induced off target mutations in the lettuce genome using high-throughput sequencing. No off-target mutations were detected at 91 homologous sites that differed by 1–5 nucleotides from the on-target site in three *BIN2*-mutated plantlets (**Supplementary Tables 1** and **2**), consistent with our findings in human cells that off-target mutations induced by CRISPR RGENs are rarely found in a single cell-derived clone[16].

Subsequently, whole plants were successfully regenerated from these genome-edited calli and grown in soil (**Fig. 2c** and **Supplementary Fig. 5**). Seeds were obtained and germinated from a fully grown homozygous biallelic mutant. As expected, the mutant allele was transmitted to the next generation (**Fig. 2d–f**). Further studies are needed to test whether the *BIN2*-disrupted lettuce has the predicted phenotype of increased BR signaling.

In conclusion, we have shown that RGEN RNPs can be used to induce targeted genome modifications in six genes in four different plant species. RGEN-induced mutations were stably maintained in whole plants that were regenerated from the protoplasts and were transmitted to the germline. Because no recombinant DNA is used in this process, the resulting genome-edited plants might be exempt from current GMO regulations[17], paving the way for the widespread use of RNA-guided genome editing in plant biotechnology and agriculture.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. NCBI Sequence Read Archive: SRX1240083.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS
This work was supported in part by grants from the Institute for Basic Science (IBS-R021-D1) and the Next-Generation BioGreen21 Program (PJ01104501 to S.C. and PJ01104502 to S.I.K.).

AUTHOR CONTRIBUTIONS
J.-S.K. and S.C. supervised the research. J.W.W., S.I.K. and C.C. carried out plant regeneration. J.K., S.W.C. H.K., S.-G.K. and S.-T.K. performed mutation analysis.

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details accompany the online version of the paper.

1. Li, J.F. *et al. Nat. Biotechnol.* **31**, 688–691 (2013).
2. Shan, Q. *et al. Nat. Biotechnol.* **31**, 686–688 (2013).
3. Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J.D. & Kamoun, S. *Nat. Biotechnol.* **31**, 691–693 (2013).
4. Kim, H. & Kim, J.S. *Nat. Rev. Genet.* **15**, 321–334 (2014).
5. Jones, H.D. *Nat. Plants* **1**, 14011 (2015).
6. Kim, S., Kim, D., Cho, S.W., Kim, J. & Kim, J.S. *Genome Res.* **24**, 1012–1019 (2014).
7. Cho, S.W., Lee, J., Carroll, D., Kim, J.S. & Lee, J. *Genetics* **195**, 1177–1180 (2013).
8. Sung, Y.H. *et al. Genome Res.* **24**, 125–131 (2014).
9. Kim, J.M., Kim, D., Kim, S. & Kim, J.S. *Nat. Commun.* **5**, 3157 (2014).
10. Kim, H.J., Lee, H.J., Kim, H., Cho, S.W. & Kim, J.S. *Genome Res.* **19**, 1279–1288 (2009).
11. Lee, H.J., Kim, E. & Kim, J.S. *Genome Res.* **20**, 81–89 (2010).
12. Bae, S., Park, J. & Kim, J.S. *Bioinformatics* **30**, 1473–1475 (2014).
13. Cho, S.W. *et al. Genome Res.* **24**, 132–141 (2014).
14. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. *Nat. Biotechnol.* **31**, 230–232 (2013).
15. Choe, S. *et al. Plant Physiol.* **130**, 1506–1515 (2002).
16. Kim, D. *et al. Nat. Methods* **12**, 237–243 (2015).
17. Kanchiswamy, C.N., Malnoy, M., Velasco, R., Kim, J.S. & Viola, R. *Trends Biotechnol.* **33**, 489–491 (2015).

ONLINE METHODS

Cas9 protein and guide RNAs. Cas9 protein tagged with a nuclear localization signal was purchased from ToolGen, Inc. (South Korea). Templates for guide RNA transcription were generated by oligo-extension using Phusion polymerase (**Supplementary Table 3**). Guide RNAs were *in vitro* transcribed through runoff reactions using the T7 RNA polymerase (New England BioLabs) according to the manufacturer's protocol. The reaction mixture was treated with DNase I (New England BioLabs) in 1× DNase I reaction buffer. Transcribed sgRNAs were resolved on an 8% denaturing urea-polyacrylamide gel with SYBR gold staining (Invitrogen) for quality control. Transcribed sgRNAs were purified with MG PCR Product Purification SV (Macrogen) and quantified by spectrometry.

Protoplast culture. Protoplasts were isolated as previously described from *Arabidopsis*[18], rice[19] and lettuce[20]. Initially, *Arabidopsis* (*Arabidopsis thaliana*) ecotype Columbia-0, rice (*Oryza sativa* L.) cv. Dongjin, and lettuce (*Lactuca sativa* L.) cv Cheongchima seeds were sterilized in a 70% ethanol, 0.4% hypochlorite solution for 15 min, washed three times in distilled water, and sown on 0.5× Murashige and Skoog solid medium supplemented with 2% sucrose. The seedlings were grown under a 16 h light (150 µmol m^{-2} s^{-1}) and 8 h dark cycle at 25 °C in a growth room. For protoplast isolation, the leaves of 14 d *Arabidopsis* seedlings, the stem and sheath of 14 d rice seedlings, and the cotyledons of 7 d lettuce seedlings were digested with enzyme solution (1.0% cellulase R10, 0.5% macerozyme R10, 0.45 M mannitol, 20 mM MES [pH 5.7], CPW solution[21]) during incubation with shaking (40 r.p.m.) for 12 h at 25 °C in darkness and then diluted with an equal volume of W5 solution[22]. The mixture was filtered before protoplasts were collected by centrifugation at 100*g* in a round-bottomed tube for 5 min. Re-suspended protoplasts were purified by floating on a CPW 21S (21% [w/v] sucrose in CPW solution, pH 5.8) followed by centrifugation at 80*g* for 7 min. The purified protoplasts were washed with W5 solution and pelleted by centrifugation at 70*g* for 5 min. Finally, protoplasts were re-suspended in W5 solution and counted under the microscope using a hemocytometer. Protoplasts were diluted to a density of 1×10^6 protoplasts/ml of MMG solution (0.4 M mannitol, 15 mM $MgCl_2$, 4 mM MES [pH 5.7]). In the case of tobacco protoplasts, 3-week-old *Nicotiana attenuata* leaves grown in B5 media were digested with enzymes (1% cellulose R10, 0.25% macerozyme R10, 0.5 M Mannitol, 8 mM $CaCl_2$, 5 mM MES [pH 5.7], 0.1% BSA) for 5 h at 25 °C in darkness. Subsequently, protoplasts were washed with an equal volume of W5 solution twice. To obtain intact protoplasts, *N. attenuata* protoplasts in W5 solution were applied to an equal volume of 21% sucrose gradient followed by swing-out centrifugation at 50*g* for 5 min. The intact protoplasts were re-suspended in W5 solution and stabilized at least for 1 h at 4 °C before PEG-mediated transfection.

Protoplast transfection. PEG-mediated RNP transfections were performed as previously described[18]. Briefly, to introduce DSBs using an RNP complex, 1×10^5 protoplast cells were transfected with Cas9 protein (10–60 µg) premixed with *in vitro*–transcribed sgRNA (20–120 µg). Prior to transfection, Cas9 protein in storage buffer (20 mM HEPES pH 7.5, 150 mM KCl, 1 mM DTT, and 10% glycerol) was mixed with sgRNA in 1× NEB buffer 3 and incubated for 10 min at room temperature. A mixture of 1×10^5 protoplasts (or 5×10^5 protoplasts in the case of lettuce) re-suspended in 200 µl MMG solution was gently mixed with 5–20 µl of RNP complex and 210 µl of freshly prepared PEG solution (40% [w/v] PEG 4000; Sigma No. 95904, 0.2 M mannitol and 0.1 M $CaCl_2$), and then incubated at 25 °C for 10 min in darkness. After incubation, 950 µL W5 solution (2 mM MES [pH 5.7], 154 mM NaCl, 125 mM $CaCl_2$ and 5 mM KCl) was added slowly. The resulting solution was mixed well by inverting the tube. Protoplasts were pelleted by centrifugation at 100*g* for 3 min and re-suspended gently in 1 ml WI solution (0.5 M mannitol, 20 mM KCl and 4 mM MES (pH 5.7)). Finally, the protoplasts were transferred into multi-well plates and cultured under dark conditions at 25 °C for 24–48 h. Cells were analyzed one day after transfection.

Protoplast regeneration. RNP-transfected protoplasts were re-suspended in 0.5× B5 culture medium[23] supplemented with 375 mg/l $CaCl_2 \bullet 2H_2O$,18.35 mg/l NaFe-EDTA, 270 mg/l sodium succinate, 103 g/l sucrose, 0.2 mg/l 2,4-dichlorophenoxyacetic acid (2,4-D), 0.3 mg/l 6-benzylaminopurine (BAP) and 0.1 g/l MES. The protoplasts were mixed with a 1:1 solution of 0.5× B5 medium and 2.4% agarose to a culture density of 2.5×10^5 protoplasts/ml. The protoplasts embedded in agarose were plated onto 6-well plates, overlaid with 2 ml of liquid 0.5× B5 culture medium, and cultured at 25 °C in darkness. After 7 days, the liquid medium was replaced with fresh culture medium. The cultures were transferred to the light (16 h light [30 µmol m^{-2} s^{-1}] and 8 h darkness) and cultured at 25 °C. After 3 weeks of culture, micro-calli grown to a few millimeters in diameter were transferred to MS regeneration medium supplemented with 30 g/l sucrose, 0.6% plant agar, 0.1 mg/l α-naphthalaneacetic acid (NAA), 0.5 mg/l BAP. Induction of multiple lettuce shoots was observed after about 4 weeks on regeneration medium.

Rooting, transfer to soil and hardening of lettuce. To regenerate whole plants, proliferated and elongated adventitious shoots were transferred to a fresh regeneration medium and incubated for 4–6 weeks at 25 °C in the light (16 h light [150 µmol m^{-2} s^{-1}] and 8 h darkness). For root induction, approximately 3–5-cm-long plantlets were excised and transferred onto a solid hormone-free 0.5× MS medium in Magenta vessels. Plantlets developed from adventitious shoots were subjected to acclimation, transplanted to potting soil, and maintained in a growth chamber at 25 °C (under cool-white fluorescent lamps with a 16-h photoperiod).

T7E1 assay. Genomic DNA was isolated from protoplasts or calli using DNeasy Plant Mini Kit (Qiagen). The target DNA region was amplified and subjected to the T7E1 assay as described previously[10]. In brief, PCR products were denatured at 95 °C and cooled down to a room temperature slowly using a thermal cycler. Annealed PCR products were incubated with T7 endonuclease I (ToolGen, Inc.) at 37 °C for 20 min and analyzed via agarose gel electrophoresis.

RGEN-RFLP. The RGEN-RFLP assay was performed as previously described[9]. Briefly, PCR products (300–400 ng) were incubated in 1× NEB buffer 3 for 60 min at 37 °C with Cas9 protein (1 µg) and sgRNA (750 ng) in a reaction volume of 10 µl. RNase A (4 µg) was then added to the reaction mixture and incubated at 37 °C for 30 min to remove the sgRNA. The reaction was stopped by adding 6× stop solution (30% glycerol, 1.2% SDS, 250 mM EDTA). DNA products were electrophoresed using a 2.5% agarose gel.

Targeted deep sequencing. The on-target and potential off-target sites were amplified from genomic DNA. Indices and sequencing adaptors were added by additional PCR. High-throughput sequencing was performed using Illumina MiSeq (v2, 300-cycle).

18. Yoo, S.D., Cho, Y.H. & Sheen, J. *Nat. Protoc.* **2**, 1565–1572 (2007).
19. Zhang, Y. *et al. Plant Methods* **7**, 30 (2011).
20. Lelivelt, C.L. *et al. Plant Mol. Biol.* **58**, 763–774 (2005).
21. Frearson, E.M., Power, J.B. & Cocking, E.C. *Dev. Biol.* **33**, 130–137 (1973).
22. Menczel, L., Nagy, F., Kiss, Z.R. & Maliga, P. *Theor. Appl. Genet.* **59**, 191–195 (1981).
23. Gamborg, O.L., Miller, R.A. & Ojima, K. *Exp. Cell Res.* **50**, 151–158 (1968).

LETTER

doi:10.1038/nature14651

Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum*

Sumiti Vinayak[1]*, Mattie C. Pawlowic[1]*, Adam Sateriale[1]*, Carrie F. Brooks[1], Caleb J. Studstill[1], Yael Bar-Peled[1], Michael J. Cipriano[1] & Boris Striepen[1,2]

Recent studies into the global causes of severe diarrhoea in young children have identified the protozoan parasite *Cryptosporidium* as the second most important diarrhoeal pathogen after rotavirus[1–3]. Diarrhoeal disease is estimated to be responsible for 10.5% of overall child mortality[4]. *Cryptosporidium* is also an opportunistic pathogen in the contexts of human immunodeficiency virus (HIV)-caused AIDS and organ transplantation[5,6]. There is no vaccine and only a single approved drug that provides no benefit for those in gravest danger: malnourished children and immunocompromised patients[7,8]. Cryptosporidiosis drug and vaccine development is limited by the poor tractability of the parasite, which includes a lack of systems for continuous culture, facile animal models, and molecular genetic tools[3,9]. Here we describe an experimental framework to genetically modify this important human pathogen. We established and optimized transfection of *C. parvum* sporozoites in tissue culture. To isolate stable transgenics we developed a mouse model that delivers sporozoites directly into the intestine, a *Cryptosporidium* clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 system, and *in vivo* selection for aminoglycoside resistance. We derived reporter parasites suitable for *in vitro* and *in vivo* drug screening, and we evaluated the basis of drug susceptibility by gene knockout. We anticipate that the ability to genetically engineer this parasite will be transformative for *Cryptosporidium* research. Genetic reporters will provide quantitative correlates for disease, cure and protection, and the role of parasite genes in these processes is now open to rigorous investigation.

Cryptosporidium infection occurs through faecal oral transmission of the environmentally resilient oocyst. The oocyst shelters four sporozoites that emerge in the small intestine and invade the epithelium. Although there is no tissue culture system for continuous passage, *C. parvum* development can be observed for 2–3 days by infecting human ileocaecal adenocarcinoma cells (HCT-8)[10]. To achieve transfection, sporozoites were excysted from oocysts purified from the faeces of experimentally infected calves using a protocol that mimics stomach and intestinal passage[11], and then electroporated before infection of HCT-8 cells (Fig. 1a). The transfection plasmids used here flanked a variety of reporter genes with candidate *C. parvum* 5′ and 3′ regulatory sequences derived from highly expressed housekeeping genes. We observed significant reporter activity 48 h after transfection using plasmids carrying nanoluciferase (Nluc; Fig. 1b), a small ATP-independent enzyme from deep sea shrimp[12], but not firefly luciferase or fluorescent proteins. Nluc luminescence correlated with the number of parasites and the amount of DNA used for transfection. Luminescence was also shown to require the presence of parasite-specific promoter elements and the introduction of DNA into parasites and not host cells (Fig. 1).

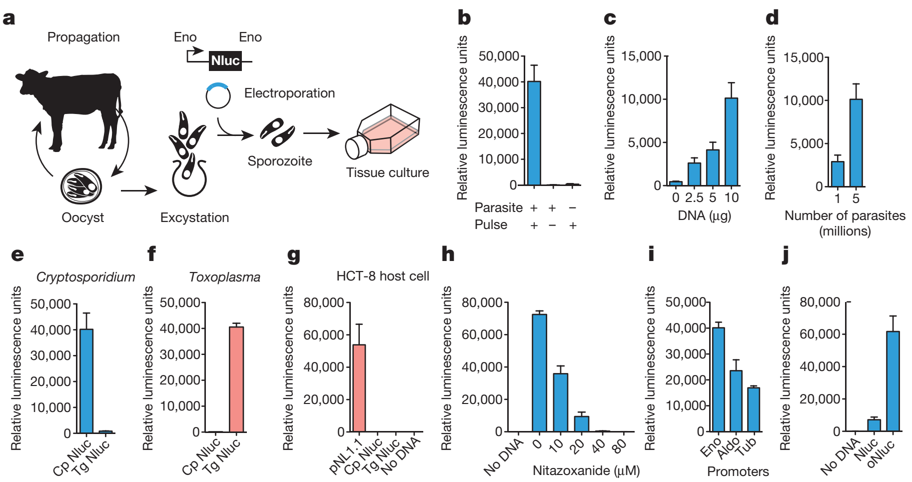


Figure 1 | Transfection of *C. parvum*. **a**, Schematic overview. *C. parvum* sporozoites were prepared from oocysts purified from infected calves and electroporated in the presence of plasmid DNA before infection of HCT-8 cells (Eno, flanking sequence from the *C. parvum* enolase gene). **b**–**j**, Luminescence measurements (the means of three technical replicates, standard deviation (s.d.) shown as error bars) of *C. parvum* (**b**–**e**, **h**–**j**, blue), *T. gondii* (**f**), or human HCT-8 cells (**g**) transfected with Nluc expression plasmids. **b**–**d**, *C. parvum* transfection requires electroporation (**b**) of DNA (**c**) into parasites (**d**). **e**, **f**, **h**, Transfection also requires plasmids to carry parasite-specific promoter sequences (**e**, **f**; testing *C. parvum* (Cp) and *T. gondii* (Tg) promoters in both parasites), and is susceptible to the *Cryptosporidium* drug nitazoxanide (**h**). **g**, Lipofection of HCT-8 cells with the original Nluc plasmid pNL1.1 (Promega), but not derived parasite vectors, results in luciferase activity in the host alone. Choice of promoter (**i**; enolase (Eno), aldolase (Aldo), α-tubulin 5′ regions (Tub) (the 3′ untranslated region (UTR) was uniformly from the enolase gene)) or codon composition (**j**; Nluc optimized to 35% GC (oNluc)) influences expression level in *C. parvum*. Note automatic gain adjustment of luminescence measurements; units are not comparable between panels. Independent biological experiments were repeated three times, and representative data are shown.

[1]Center for Tropical and Emerging Global Diseases, University of Georgia, Paul D. Coverdell Center, 500 D.W. Brooks Drive, Athens, Georgia 30602, USA. [2]Department of Cellular Biology, University of Georgia, Paul D. Coverdell Center, 500 D.W. Brooks Drive, Athens, Georgia 30602, USA.
*These authors contributed equally to this work.

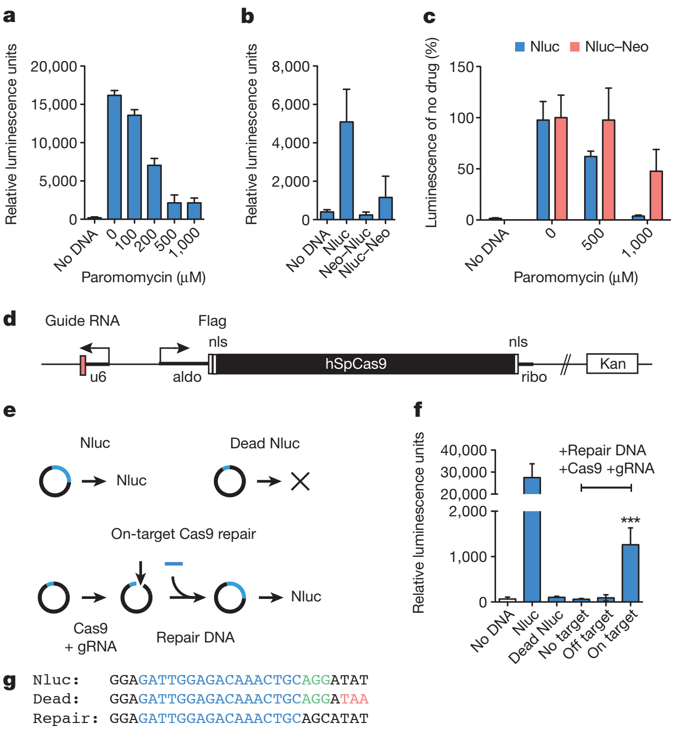


Figure 2 | Luciferase assays for *C. parvum* drug resistance and CRISPR/Cas9 activity. **a**, HCT-8 cells were infected with Nluc-transfected sporozoites and grown for 2 days in the presence of paromomycin. **b**, Translational fusions were constructed placing Neo at the amino or carboxy terminus of Nluc. Nluc–Neo shows luciferase activity, albeit at a reduced level when compared to Nluc alone. **c**, *C. parvum* transfected with Nluc (blue) or Nluc–Neo (red) were grown in different concentrations of paromomycin. Luciferase activity for each plasmid was normalized to its drug-free level. **d**, CRISPR/Cas9 plasmid for *C. parvum*. Flag, epitope tag; nls, nuclear localization signal; ribo, ribosomal protein L13A 3′ UTR; u6, newly annotated promoter CM000433:553110–553472. **e**, **g**, Outline (**e**) and sequences (**g**) for Nluc repair assay. Guide RNA target, blue; protospacer adjacent motif, green; mutagenized codon 18, red. **f**, Sporozoites were transfected with Nluc or a codon 18 termination mutant (Dead Nluc); note ablation of signal. In addition to the Dead Nluc plasmid, some parasites also received a 125 bp double-stranded repair DNA fragment, and the Cas9 plasmid with the indicated guide RNAs (gRNAs; no target, empty gRNA cassette; off target, GFP gRNA; on target, Nluc gRNA). Statistical analysis compares Dead Nluc alone with Dead Nluc and Cas9 and specific gRNA. Note significant Cas9-mediated restoration of luciferase activity (***$P = 0.0006$, unpaired *t*-test). $n = 3$ technical replicates for **a**–**c**, and controls from **f**; $n = 6$ technical replicates for on-target samples in **f**. Error bars are s.d. and all experiments depicted here were repeated three times and representative data are shown.

Furthermore, reporter signal was ablated by the anti-parasitic drug nitazoxanide. Transient transfection of *C. parvum* is inefficient (<10,000 fold when compared to the related apicomplexan *Toxoplasma gondii* in parallel experiments) and requires a highly sensitive reporter such as Nluc to be noticeable.

In an effort to enhance efficiency we evaluated different electroporation devices, electrical wave programs and buffer compositions (Extended Data Fig. 1); this produced tenfold enhancement. We tested flanking sequences from different *C. parvum* genes and identified the enolase promoter to be strongest. The *C. parvum* genome is AT rich and shows strong codon bias[13]. We also noted a preference for A over T within the first 20 codons and thus explored codon optimization and found sixfold enhancement (Fig. 1j).

To enable enrichment of transgenic parasites, we next explored the selection of drug resistance. The aminoglycoside antibiotic paromomycin does not cure cryptosporidiosis in people, but is effective in tissue culture (Fig. 2a) and in immunocompromised mice[14]. Work in other protist models has shown aminoglycoside phosphotransferases to

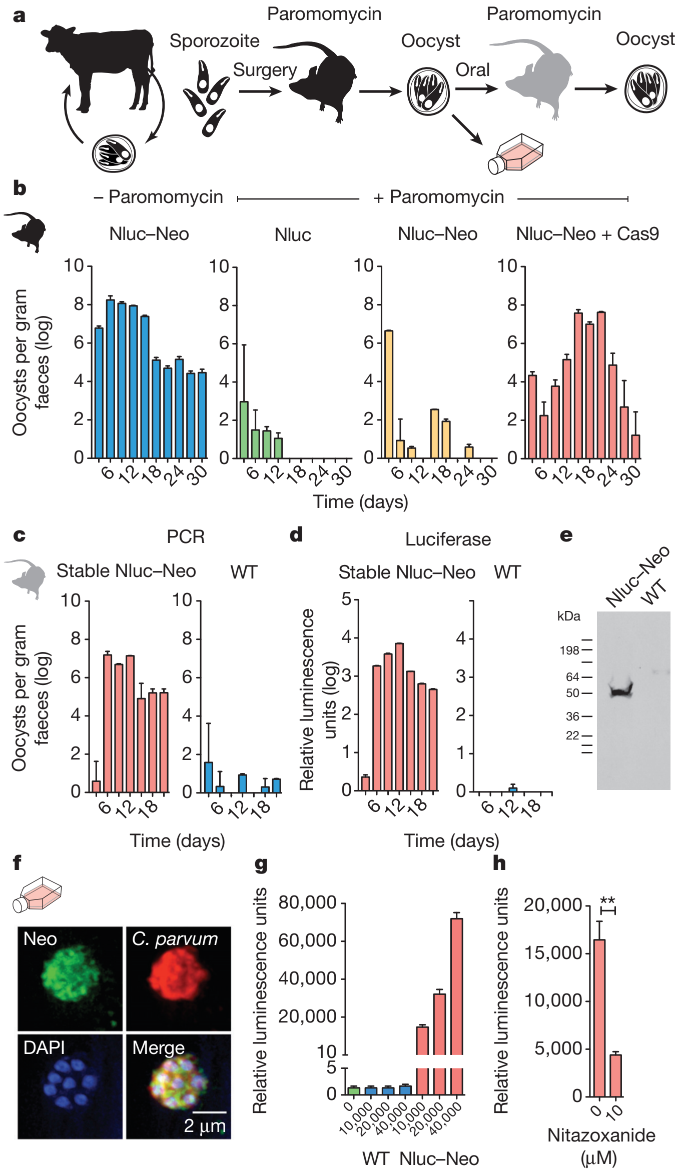


Figure 3 | Mouse model for selection of stable *C. parvum* transgenics. **a**, Outline of the selection strategy. Transfected sporozoites were injected into the small intestine by surgery (Extended Data Fig. 2) and mice were treated with paromomycin. Oocysts were purified from the faeces and used to infect cultures or mice by oral gavage. **b**, Quantitative PCR of *C. parvum* DNA isolated from faeces of mice infected with transfected sporozoites (four mice per group) and treated as indicated. Emergence of paromomycin resistance required the Nluc–Neo and Cas9 plasmids. **c**, **d**, Upon reinfection, parasites show strong drug resistance (**c**) and luciferase activity (**d**). In repeat experiments we noted that luciferase is detectable as early as 6 days after transfection in the faeces of the first infected mouse (Extended Data Fig. 4). **e**, Protein extracts from oocysts were analysed by SDS-polyacrylamide gel electrophoresis (SDS–PAGE) and western blot using an antibody against Neo (rabbit anti-neomycin phosphotransferase II; EMD Millipore). Predicted molecular mass of the Nluc–Neo fusion protein is 48.3 kDa. **f**, Immunofluorescence staining using anti-Neo (mouse anti-Neo; Alpha Diagnostic International) and *C. parvum* (tryptophan synthase B) antibodies. Note multiple nuclei in 4′,6-diamidino-2-phenylindole (DAPI) stain typical for *C. parvum* meronts. No anti-Neo staining was observed in wild-type parasites. **g**, Luciferase assays for HCT-8 cultures infected with wild-type (WT; blue) and transgenic (Nluc–Neo; red) parasites. The *y*-axis is split to show level of luminescence background. $n = 3$ technical replicates, error bars are s.d., the experiment was done twice. **h**, Ninety-six-well luciferase drug assay using 1,000 oocysts per well. Note significant growth inhibition on treatment with 10 μM nitazoxanide (**$P = 0.0036$, unpaired *t*-test). $n = 3$ technical replicates, error bars are s.d., the experiment was repeated two times and representative data are shown.

confer resistance to paromomycin[15,16]. Appreciation of *C. parvum* drug resistance in culture is complicated by the lack of continuous growth. We thus constructed translational fusions between the Nluc reporter and the neomycin resistance marker (Neo)[15] to focus our observation on the small subset of transfected parasites. Luciferase activity in parasites expressing Nluc–Neo showed reduced susceptibility to paromomycin treatment compared to Nluc alone (Fig. 2c), and thus we concluded that Nluc–Neo confers drug resistance in this transient assay.

Our genome searches indicated that *Cryptosporidium* species lack non-homologous end joining DNA repair. This suggested transgene integration to be rare and to require homologous recombination[17,18]. Such recombination can be enhanced by long flanking regions and/or double-strand breaks introduced by restriction enzymes, transcription activator-like effector nucleases (TALENs) or CRISPR/Cas9 (refs 18, 19). To build a *C. parvum* CRISPR/Cas9 system, we constructed a plasmid in which the *C. parvum* U6 RNA promoter drives a guide RNA cassette[20] and the *Streptococcus pyogenes Cas9* gene[21] is flanked by parasite regulatory sequences (Fig. 2d). To test this system, we conducted a Cas9-dependent DNA repair experiment (Fig. 2e–g). We introduced a stop codon into the Nluc reporter that ablated luciferase activity (Dead Nluc). We then targeted the dead gene with a guide RNA, and provided a short double-stranded template for repair that restores read-through translation and renders the repaired gene resistant to further Cas9 cutting. When *C. parvum* sporozoites are co-transfected with a specific guide, luciferase activity is restored ($P = 0.0006$, unpaired t-test). No change is observed with no or off-target guides.

Interferon-γ knockout mice are susceptible to *C. parvum* infection through oral inoculation of oocysts[22]. However, infection with free sporozoites is less effective[23], probably due to stomach passage. We developed a surgical protocol to inject transfected sporozoites directly into the small intestine to maximize infection (Extended Data Fig. 2). When mice were killed 24 h after infection, luciferase activity was observed in scrapings of the intestinal epithelium. We also established an effective treatment protocol using paromomycin supplementation of the drinking water (Extended Data Fig. 3).

Next, we infected mice by surgery with transfected sporozoites and treated them with paromomycin as indicated (Fig. 3 and Extended Data Fig. 4; four mice per group). Faeces were collected every 3 days and oocyst shedding was measured by quantitative polymerase chain reaction (PCR) targeting the *C. parvum* 18S ribosomal RNA locus. Mice infected with parasites transfected with the Nluc–Neo plasmid that did not receive drug shed high numbers of oocysts and remained infected for the 30 days observed (Fig. 3b, blue). Those infected with parasites that received the Nluc plasmid (lacking the Neo gene; Fig. 3b, green) were rapidly cured by drug treatment. Those transfected with Nluc–Neo alone and drug treated were also cured (infection may persist slightly longer). In contrast, infection with parasites carrying the Nluc–Neo plasmid and the Cas9 plasmid (Fig. 3b, red; Cas9 target detailed later) rapidly rebounded to levels similar to untreated mice. Oocysts emerging from selection were purified from faeces and used to infect mice that were again treated with paromomycin; wild-type oocysts were used in parallel (100,000 oocysts per mouse by gavage). While paromomycin treatment cured infection with wild-type parasites, transgenic parasites showed immediate robust drug resistance (Fig. 3c). When these oocysts were probed by western blot with anti-Neo antibody, we detected a band consistent with an Nluc–Neo fusion protein.

Purified oocysts were also used to infect cell cultures, and processed for immunofluorescence after 2 days. Transgenic but not wild-type intracellular parasite stages showed fluorescence when probed with antibodies specific for either Neo or Nluc (Fig. 3f and data not shown). These cultures also displayed strong luciferase activity not observed in wild type. This activity exceeded that previously observed in transient transfection experiments by five orders of magnitude on a per-cell basis. We assessed whether these organisms could be suitable for drug-screening assays by infecting 96-well plates with 1,000 oocysts per well and measured luciferase after 48 h. Infected wells were clearly distinguishable from uninfected wells ($z' > 0.6$; $n = 20$). Similarly, wells treated with nitazoxanide showed significant growth inhibition ($P = 0.0036$, unpaired t-test). Luciferase also provided a convenient way to assess the infection state of animals. We sampled 10 mg of faeces from mice diagnosed in parallel by PCR and found this assay to be sensitive, specific and faster than PCR (Fig. 3d). We note that Nluc expression remains stable when parasites are propagated in mice in the absence of paromomycin (Extended Data Fig. 5).

Cryptosporidium is remarkably resistant to antifolates, a mainstay of treatment against other apicomplexans, and this resistance has been attributed to differences in the target enzyme dihydrofolate reductase-thymidylate synthase (DHFR-TS)[24]. However, *Cryptosporidium* is

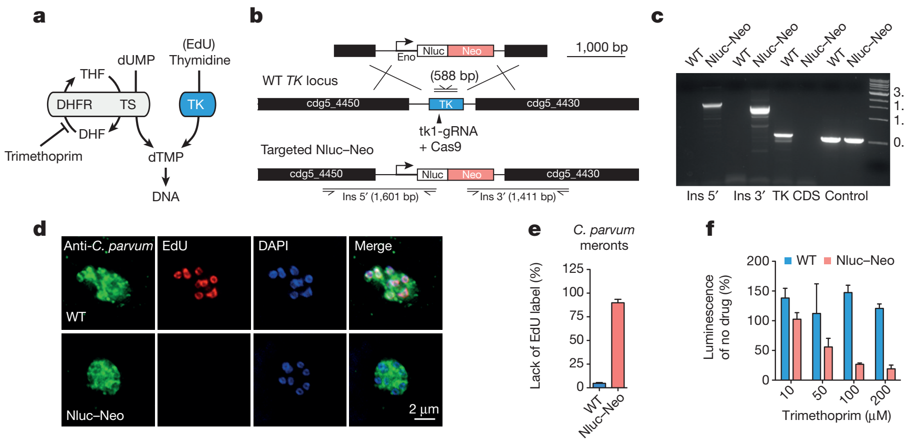


Figure 4 | Targeted deletion of *C. parvum* TK. **a**, Owing to a horizontal gene transfer, *C. parvum* has two pathways to synthesize dTMP: TK and DHFR-TS. DHF, dihydrofolic acid; THF, tetrahydrofolic acid; dUMP, uridine monophosphate. **b**, Map of the *C. parvum TK* locus, the targeting plasmid and the predicted modified locus. Primers and amplicon sizes of diagnostic PCR products are indicated (Ins, insertion). **c**, PCR analysis using genomic DNA from wild-type (WT) and transgenic parasites (Nluc–Neo, oocysts purified from faeces of infected mice shown in Fig. 3c; CDS, coding sequence). Primer sequences are provided in Supplementary Table 1. **e**, Quantification of EdU-labelling experiments (meronts with four or more nuclei were scored, two biological repeats, $n = 105$ each sample, error bars are s.d.), **d**, Representative fluorescence micrographs are shown. Antibody to *C. parvum* tryptophan synthase B was used to identify parasites (green). **f**, Trimethoprim treatment of wild-type (blue) and Nluc–Neo transgenic (red) parasites. Wild-type parasites were measured in transient transfection assays with Nluc plasmid ($n = 3$, technical replicates, error bars are s.d.). The assay shown was conducted in the presence of 10 μM thymidine to avoid indirect host cell toxicity[30] (experiments without thymidine produced indistinguishable results). Experiments were repeated three times and representative data are shown.

unique among apicomplexans in that it acquired a thymidine kinase (TK) by horizontal gene transfer from bacteria[25]. We hypothesized that TK may also contribute to *Cryptosporidium* antifolate resistance by providing an alternative route to thymidine monophosphate (dTMP; Fig. 4a). For this reason, the TK locus was targeted for insertion, allowing us to test this hypothesis by gene disruption. We mapped the locus in stable transgenic parasites by PCR using primers that link the marker genes with genomic sequences beyond the flanking regions on the targeting construct. This mapping is consistent with insertion by homologous double crossover (Fig. 4b, c). Furthermore, the TK coding sequence is no longer detectable, indicating uniform loss of the gene in the selected population. We tested for DNA incorporation of the thymidine analogue 5-ethynyl-2′-deoxyuridine (EdU) using click chemistry and fluorescence microscopy[26]. Wild-type parasites grown in the presence of EdU show fluorescent nuclei. This labelling is lost in the transgenic parasites (Fig. 4d, e), confirming loss of TK at the biochemical level. We next treated parasite infected cultures with the antifolate trimethoprim. We confirmed the previously observed resistance in wild-type parasites, but noted enhanced susceptibility in the mutants (Fig. 4f). We conclude that the *C. parvum* TK is a non-essential enzyme required for the activation of thymidine, and that its presence limits the efficacy of antifolate therapy in *Cryptosporidium*.

We show that major hurdles towards genetic analysis and manipulation for cryptosporidiosis can be overcome by maximizing the efficiency of each step of the process and by focusing on *in vivo* propagation and selection. There is an urgent need for new anti-parasitic drugs[3]. *Cryptosporidium* is not susceptible to drugs widely used against related pathogens, which reflects substantial differences in its metabolism and metabolite uptake[27]. Luciferase reporter parasites enable phenotypic screening in culture and animals with sufficient sensitivity and specificity to warrant a comprehensive effort to discover novel compounds. Gene deletion now permits biological target validation. Genetic modification may also allow the construction of attenuated parasites as a potential oral vaccine. While infants and toddlers are highly susceptible to the disease, infection is rarely detected in older children[1,3]. This is consistent with infection studies in people and animals suggesting the development of anti-parasitic and anti-disease immunity[3,28,29]. A better understanding of the mechanisms underlying disease and protection will be required to design and produce such a vaccine.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 May; accepted 12 June 2015.
Published online 15 July 2015.

1. Kotloff, K. L. *et al.* Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* **382,** 209–222 (2013).
2. Mondal, D. *et al.* Contribution of enteric infection, altered intestinal barrier function, and maternal malnutrition to infant malnutrition in Bangladesh. *Clin. Infect. Dis.* **54,** 185–192 (2012).
3. Checkley, W. *et al.* A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for cryptosporidium. *Lancet Infect. Dis.* **15,** 85–94 (2015).
4. Liu, L. *et al.* Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet* **379,** 2151–2161 (2012).
5. Raja, K. *et al.* Prevalence of cryptosporidiosis in renal transplant recipients presenting with acute diarrhea at a single center in Pakistan. *J. Nephropathol.* **3,** 127–131 (2014).
6. Hunter, P. R. & Nichols, G. Epidemiology and clinical features of *Cryptosporidium* infection in immunocompromised patients. *Clin. Microbiol. Rev.* **15,** 145–154 (2002).
7. Amadi, B. *et al.* Effect of nitazoxanide on morbidity and mortality in Zambian children with cryptosporidiosis: a randomised controlled trial. *Lancet* **360,** 1375–1380 (2002).
8. Amadi, B. *et al.* High dose prolonged treatment with nitazoxanide is not effective for cryptosporidiosis in HIV positive Zambian children: a randomised controlled trial. *BMC Infect. Dis.* **9,** 195 (2009).
9. Striepen, B. Parasitic infections: time to tackle cryptosporidiosis. *Nature* **503,** 189–191 (2013).
10. Upton, S. J., Tilley, M. & Brillhart, D. B. Comparative development of *Cryptosporidium parvum* (Apicomplexa) in 11 continuous host cell lines. *FEMS Microbiol. Lett.* **118,** 233–236 (1994).
11. Gut, J. & Nelson, R. G. *Cryptosporidium parvum*: synchronized excystation *in vitro* and evaluation of sporozoite infectivity with a new lectin-based assay. *J. Eukaryot. Microbiol.* **46,** 56S–57S (1999).
12. Hall, M. P. *et al.* Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem. Biol.* **7,** 1848–1857 (2012).
13. Abrahamsen, M. S. *et al.* Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304,** 441–445 (2004).
14. Theodos, C. M., Griffiths, J. K., D'Onfro, J., Fairfield, A. & Tzipori, S. Efficacy of nitazoxanide against *Cryptosporidium parvum* in cell culture and in animal models. *Antimicrob. Agents Chemother.* **42,** 1959–1965 (1998).
15. Mochizuki, K. High efficiency transformation of *Tetrahymena* using a codon-optimized neomycin resistance gene. *Gene* **425,** 79–83 (2008).
16. Gueiros-Filho, F. J. & Beverley, S. M. On the introduction of genetically-modified *Leishmania* outside the laboratory. *Exp. Parasitol.* **78,** 425–428 (1994).
17. Fox, B. A., Ristuccia, J. G., Gigley, J. P. & Bzik, D. J. Efficient gene replacements in *Toxoplasma gondii* strains deficient for nonhomologous end-joining. *Eukaryot. Cell* **8,** 520–529 (2009).
18. Lee, A. H., Symington, L. S. & Fidock, D. A. DNA repair mechanisms and their biological roles in the malaria parasite *Plasmodium falciparum*. *Microbiol. Mol. Biol. Rev.* **78,** 469–486 (2014).
19. Brooks, C. F. *et al.* The *Toxoplasma* apicoplast phosphate translocator links cytosolic and apicoplast metabolism and is essential for parasite survival. *Cell Host Microbe* **7,** 62–73 (2010).
20. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
21. Sidik, S. M., Hackett, C. G., Tran, F., Westwood, N. J. & Lourido, S. Efficient genome engineering of *Toxoplasma gondii* using CRISPR/Cas9. *PLoS ONE* **9,** e100450 (2014).
22. Griffiths, J. K., Theodos, C., Paris, M. & Tzipori, S. The gamma interferon gene knockout mouse: a highly sensitive model for evaluation of therapeutic agents against *Cryptosporidium parvum*. *J. Clin. Microbiol.* **36,** 2503–2508 (1998).
23. Fayer, R., Nerad, T., Rall, W., Lindsay, D. S. & Blagburn, B. L. Studies on cryopreservation of *Cryptosporidium parvum*. *J. Parasitol.* **77,** 357–361 (1991).
24. Liu, J., Bolstad, D. B., Bolstad, E. S. D., Wright, D. L. & Anderson, A. C. Towards new antifolates targeting eukaryotic opportunistic infections. *Eukaryot. Cell* **8,** 483–486 (2009).
25. Striepen, B. *et al.* Gene transfer in the evolution of parasite nucleotide biosynthesis. *Proc. Natl Acad. Sci. USA* **101,** 3154–3159 (2004).
26. Salic, A. & Mitchison, T. J. A chemical method for fast and sensitive detection of DNA synthesis *in vivo*. *Proc. Natl Acad. Sci. USA* **105,** 2415–2420 (2008).
27. Striepen, B. in *Antimicrobial Drug Resistance* Vol. 1 (eds Mayers, D. L., Lerner, S. A., Ouellette, M. & Sobel, J. D.) 605–621 (Springer, 2009).
28. Sheoran, A., Wiffin, A., Widmer, G., Singh, P. & Tzipori, S. Infection with *Cryptosporidium hominis* provides incomplete protection of the host against *Cryptosporidium parvum*. *J. Infect. Dis.* **205,** 1019–1023 (2012).
29. McDonald, V., Deer, R., Uni, S., Iseki, M. & Bancroft, G. J. Immune responses to *Cryptosporidium muris* and *Cryptosporidium parvum* in adult immunocompetent or immunocompromised (nude and SCID) mice. *Infect. Immun.* **60,** 3325–3331 (1992).
30. Jiang, L., Lee, P. C., White, J. & Rathod, P. K. Potent and selective activity of a combination of thymidine and 1843U89, a folate-based thymidylate synthase inhibitor, against *Plasmodium falciparum*. *Antimicrob. Agents Chemother.* **44,** 1047–1050 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Sharling for initial contributions and L. Hedstrom, J. Mead, S. Vaishnava, L. Xiao and Y. Belkaid for discussion. This work was funded in part by the National Institutes of Health (NIH; R01AI112427) to B.S. and by a pilot grant from the Centers for Disease Control and the University of Georgia Research Foundation to B.S. and L. Xiao. M.J.C. was supported by training grant NIH T32AI060546 and B.S. is a Georgia Research Alliance Distinguished Investigator.

Author Contributions S.V. developed the transfection and luciferase assay; M.C.P. optimized transfection and developed the Cas9 system; S.V., M.C.P., A.S. and C.F.B. developed the mouse infection protocol and A.S. developed selection assays; C.F.B. developed surgery; C.J.S. and Y.B.-P. constructed some of the plasmids; and M.J.C. provided bioinformatics support. S.V., M.C.P., A.S. and C.F.B. conducted animal experiments and genotypic and phenotypic characterization. S.V., M.C.P., A.S., C.F.B. and B.S. conceived the study and B.S. wrote the manuscript with contributions from S.V., M.C.P. and A.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.S. (striepen@uga.edu).

METHODS

***C. parvum* reporter and drug resistance vectors.** *C. parvum* transfection vectors were derived from plasmid pH_3BG[19] and modified to contain *C. parvum* promoter and 5′ and 3′ untranslated messenger RNA regions. We mined the genome and a variety of expression data sets collectively available through Crypto DB (http://www.cryptoDB.org)[31] to identify genes that are highly expressed across the life-cycle. Promoters and 5′ UTRs of the enolase (cgd5_1960), α-tubulin (cgd4_2860), and aldolase (cgd1_3020) genes and 3′ UTRs of enolase (51 bp), α-tubulin (97 bp) or ribosomal protein L13A (cgd5_970, UTR 211 bp) were amplified from genomic DNA by PCR (see Supplementary Table 1 for a list of primer sequences and restriction sites used). Nluc was amplified from pNL1.1 (Promega Corporation), firefly luciferase and different fluorescent protein genes were amplified from vectors used for *T. gondii*[32–34]. The neomycin resistance gene was amplified from plasmid pNeo4 (ref. 15) (a gift from J. Gaertig, University of Georgia) and introduced 5′ or 3′ of Nluc in a plasmid with enolase regulatory sequences. To target the *TK* gene, regions flanking the gene were amplified and introduced into the Nluc–Neo vector (the promoter but not the 3′ UTR was retained).

***C. parvum* CRISPR/Cas9 genome editing.** Human codon-optimized *Streptococcus pyogenes* Cas9 (hSpCas9) carrying a Flag tag and N- and C-terminal nuclear localization signals was amplified from pX330 (ref. 35) and introduced into the Aldolase-Nluc-ribo vector replacing the Nluc. A guide RNA cassette was synthesized containing the *C. parvum* U6 promoter identified by genome searches using known structural RNA sequences from *Plasmodium falciparum*[36], two inverted BbsI restriction sites to facilitate guide cloning, a *trans*-activating CRISPR RNA (tracrRNA) consensus sequence and a terminator (poly T) sequence, and was introduced into the Cas9 plasmid.

To test for CRISPR/Cas9-mediated repair *in vitro*, we modified the codon-optimized Nluc vector by introducing a premature stop codon (Y18Stop) adjacent to a guide target sequence at the beginning of the gene by site-directed mutagenesis (QuikChange II, Agilent Technologies). A 125 bp double-stranded (ds)DNA oligonucleotide was synthesized that restored Y18 and disrupted the PAM motif (G17A) of the guide RNA target, thus rendering it resistant to further Cas9 cuts.

Parasite excystation and transfection. Oocyst excystation was carried out as described[11] with some modification. Up to 10^8 *C. parvum* Iowa strain oocysts (Sterling Parasitology Laboratory or Bunch Grass Farm) were suspended in 100 μl of 1:4 aqueous dilution of 5.25% sodium hypochlorite and incubated on ice for 5 min. Oocysts were then washed three times with ice-cold PBS, suspended at 3.9×10^5 oocysts per ml of 0.2 mM sodium taurocholate (prepared in PBS) and incubated at 15 °C (10 min) and then at 37 °C (60–90 min). Emergence of sporozoites was monitored microscopically (typical efficiency 70–90%). Sporozoites were filtered through a 3 μM polycarbonate filter to remove unexcysted oocysts, washed with ice-cold PBS, and counted.

Initially we used a BTX ECM 630 device for electroporation (Harvard Apparatus). Excysted sporozoites (10^7) were suspended in complete cytomix buffer (120 mM KCl, 0.15 mM $CaCl_2$, 10 mM K_2HPO_4/KH_2PO4, pH 7.6, 25 mM HEPES, pH 7.6, 2 mM EGTA, 5 mM $MgCl_2$, pH 7.6, supplemented with 2 mM ATP and 5 mM glutathione), mixed with plasmid DNA, and electroporated with a single 1,500 V pulse, resistance of 25 Ω, and a capacitance of 25 μF. To enhance transfection efficiency, we switched to using the AMAXA Nucleofactor 4D device (Lonza Cologne GmbH). After excystation, 10^7 sporozoites were suspended in 15 μl Lonza SF Buffer and combined with 10–50 μg DNA (prepared in Tris-EDTA, pH 8.0) at a final volume of 20 μl. The parasite–DNA mix was added to small, strip cuvettes and electroporated using program EH100. Additional electroporation conditions were explored to arrive at this protocol and those are listed in Extended Data Fig. 1.

For *in vitro* transfection assays, human ileocaecal adenocarcinoma (HCT-8) cells (ATCC) were grown in RPMI-1640 with glutamine supplemented with 10% FBS, 1 mM sodium pyruvate, 50 U ml^{-1} penicillin, 50 μg ml^{-1} streptomycin and amphotericin B in 24-, 48- or 96-well plates to 70% confluency. No effort was made to authenticate this cell line or test for mycoplasma. Prior to infection, media was replaced with DMEM with 2% FBS, 50 U ml^{-1} penicillin, 50 μg ml^{-1} streptomycin and amphotericin B, and 0.2 mM L-glutamine. For *in vivo* experiments electroporated sporozoites were suspended in PBS and kept on ice until administered to the mice.

The *T. gondii* Nluc plasmid was constructed by inserting the Nluc sequence into vector $pCTH_3$ (ref. 32) and parasites were electroporated and used to infect human foreskin fibroblasts as described[37]. HCT-8 cells were cultured in 24-well plates until confluent, transfected with 500 ng of DNA using Lipofectamine 2000 as described by the manufacturer (Life Technologies), and assayed for Nluc activity after 48 h.

Animal ethics statement. Animal experiments were approved by the Institutional Animal Care and Use Committee of the University of Georgia (animal use protocol no. A2012 03-028-Y3-A12).

Surgical delivery of transfected sporozites into IFN-γ-deficient mice. In preliminary experiments we noted that antibiotic removal of bacterial flora enhances susceptibility of mice. Prior to infection mice were orally treated by gavage daily for a week before infection with an antibiotic cocktail (3 mg ampicillin, 3 mg streptomycin, 0.95 mg metronidazole, 3 mg neomycin and 1.5 mg vancomycin in distilled H_2O, per mouse/per day; all antibiotics purchased from Sigma). To deliver sporozoites directly to the small intestine, we developed a mouse survival surgery protocol for female C57BL/6 IFN-γ-deficient mice (B6.129S7-Ifng^{tm1Ts}/J, Jackson Laboratories) aged 6–8 weeks. The abdominal area of mice was shaved with clippers. Animals were placed in an isofluorane (3–5%) anaesthesia induction chamber and then moved to a nosecone (1–3% isofluorane as needed) on a sterile surgical field. A sterile drape was applied over a warming pad after sterilization of the area with 70% ethanol. Respiration and response to stimulation (toe pinch) were monitored during the procedure and the vaporizer adjusted as needed. Mucous membranes and footpads were monitored for colour to confirm adequate perfusion. Three betadine (Povidone-iodine) scrubs followed by a 70% ethanol wipe were applied to shaved skin before surgery. Ophthalmic ointment (Puralube, Dechra Veterinary Products) was applied to prevent drying of eyes. Skin was vertically incised midline of the abdominal region below the sternum with microsurgical scissors for approximately 1.5 cm followed by vertical incision of the peritoneum. Exposed jejunum/ileum was injected with 10^7 transfected sporozoites suspended in 200 μl PBS containing sterile food colouring dye as tracer. After injection, suturing was performed to close the peritoneum. Mice were administered 0.01–0.02 ml per gram body weight of warm lactated Ringer's solution subcutaneously after surgery. Meloxicam analgesic was also administered to the mice after surgery. At completion of the procedure, the eye ointment was wiped off and the vaporizer was turned off and the mice were allowed to breathe the oxygen supply gas until they began to wake. Mice were placed in a recovery area until ambulatory and exhibiting normal respiration and were watched for 2 h after surgery. Incision sites were monitored daily until fully healed (10–14 days). Twenty-four hours after surgical infection, water in mouse cages was replaced with distilled H_2O containing 16 mg ml^{-1} paromomycin, a concentration we determined to deliver a daily dose of 40 mg kg^{-1} paromomycin to each mouse (Extended Data Fig. 3). Mice were randomly assigned to groups before surgery. A sample size of four animals per treatment group was judged to be sufficiently large enough to draw appropriate conclusions. All mice survived surgery and were included in the results reported here. Investigators were not blinded to group allocation during the experiments.

Mouse faeces collection and storage. Faecal samples were collected from mice (typically four mice per cage) starting 3 days after infection every third day for up to a month. Mice were transferred to a fresh, sterile cage for 2–3 h, and faeces from the cage were collected, pooled, and stored at 4 °C.

Luciferase assay. For transient transfection experiments, electroporated sporozoites were added to 70% confluent HCT-8 culture and infection was allowed to proceed at 37 °C for 48 h. Media was removed from wells and 200 μl of NanoGlo lysis buffer supplemented with NanoGlo substrate (1:50, Promega Corporation) was added to each well. Cells were scraped and the lysate was transferred to white 96-well plates and luminescence was measured using a Synergy H4 Hybrid Microplate Reader (BioTek Instruments). For drug assays with purified stable transgenic oocysts, the culture supernatant was collected after 48 h from 96-well plates. An equal amount of supernatant and NanoGlo lysis buffer with substrate was combined and luminescence was measured.

For luciferase measurement from mouse faecal samples, 20 mg of faeces was weighed into a 1.5-ml microcentrifuge tube and homogenized in 1 ml of lysis buffer (50 mM Tris-HCl, 10% glycerol, 1% Triton-X, 2 mM dithiothreitol (DTT), 2 mM EDTA) using 10–15 glass beads (3 mm) and a vortex mixer for 1 min, followed by clarification of lysate by brief centrifugation. One-hundred microlitres of lysate was mixed with an equal volume of NanoGlo Luciferase Buffer (prepared with 1:50 dilution of substrate) and luminescence was measured as described.

High-throughput imaging assay for parasite growth. For drug assays we used either luciferase activity or a 96-well infection and imaging protocol[38] using a BD Pathway instrument. Parasites and host cells were quantified using an ImageJ macro adapted from ref. 39. The ratio of parasites to host nuclei was determined for each sample image and normalized to untreated controls.

For oocyst quantification by high-throughput microscopy, we weighed collected mouse faeces and diluted in PBS (5 μl mg^{-1}). Samples were incubated at 95 °C for 10 min, vortexing every 2 min at high speed. Large debris was allowed to settle for 10 min, then 10 μl of the suspension were mixed with 990 μl PBS and 1 μl of fluorescein isothiocyanate (FITC)-conjugated goat polyclonal anti-*Cryptosporidium* antibody (GeneTex). After 1 h at room temperature, the sample was centrifuged at 2,000*g* for 15 min. The pellet was suspended in 200 μl PBS and transferred to a 96-well plate for microscopy. Plates were imaged using BD Pathway and oocysts were counted

using an ImageJ macro. Using a standard curve (uninfected mouse faeces spiked with known amounts of oocysts), oocyst counts were converted to oocysts per grams faeces.

Quantification of oocyst shedding using qRT–PCR. DNA was extracted from 100 mg faeces using ZR Faecal DNA MiniPrep Kit (Zymo Research Corporation) following the manufacturer's protocol with slight modification. While in lysis buffer, the sample was freeze-thawed in liquid nitrogen five times before the first centrifugation step. Each sample was eluted in 50 μl water, 1 μl of eluate was used for qRT–PCR along with 10 μM primers targeting *Cryptosporidium* 18S rRNA[40] and SYBR Master Mix (Life Technologies) for detection. Each qRT–PCR reaction was normalized using an eight-point standard curve (faecal DNA purified from uninfected mouse faeces spiked with known amounts of oocysts) for each set of samples.

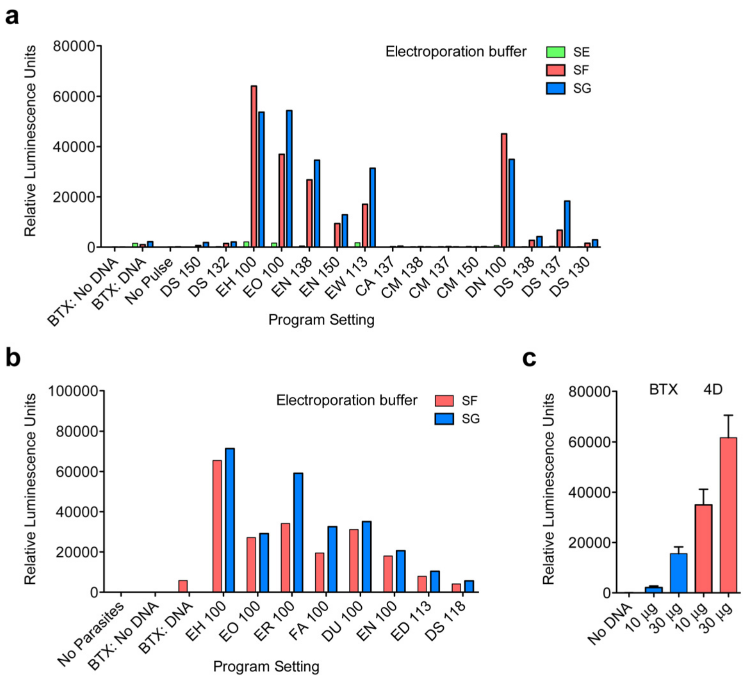
Oocyst purification from mouse faeces. Oocysts were purified from faeces using sucrose suspension followed by a caesium chloride centrifugation[41]. Mouse faeces were suspended in tap water, passed through a 850-μm mesh filter, followed by 250-μm mesh. This filtered suspension was mixed 1:1 with aqueous sucrose solution (specific gravity 1.33), and centrifuged at 1,000*g* for 5 min. Oocysts were collected from the supernatant and suspended in 0.85% saline solution. 0.5 ml of this preparation was overlaid onto 0.8 ml of 1.15 specific gravity CsCl, and centrifuged for 3 min at 16,000*g*. Oocysts were collected from the top ml of the sample, washed in 0.85% saline, counted with disposable counting chamber (KOVA International) and suspended in 2.5% potassium dichromate for storage at 4 °C.

Western blotting. For western blot analysis, oocysts from wild-type and transgenic Nluc–Neo parasites were excysted as described earlier and sporozoites were lysed in SDS sample buffer. Protein extract from 10^7 sporozoites was loaded per lane and subjected to electrophoresis on a precast Any kD Mini-PROTEAN TGX gel (Bio-Rad) followed by transfer to 0.2-μm nitrocellulose membrane (Bio-Rad). Blots were blocked and probed with an anti-neomycin phosphotransferase II antibody (EMD Millipore) at 1:1,000 dilution and goat anti-rabbit IgG (H + L)–HRP conjugate (Bio-Rad) at 1:20,000 dilution followed by detection with ECL Western Blotting Substrate (Thermo Pierce) and exposure to film. Equal loading of blots was controlled by stripping and reprobing with an antibody to α-tubulin.

EdU labelling and immunofluorescence microscopy. EdU labelling was performed using the Click-iT EdU Alexa Fluor 594 Imaging Kit following the manufacturer's instructions (Life Technologies). Purified stable transgenic oocysts expressing the luciferase or wild-type oocysts were inoculated into 24-well plates containing coverslips confluent with HCT-8 cells. After 24 h, EdU was added to the media at 10 μM and left for 18 h before fixation. For immunofluorescence, primary antibodies used were mouse monoclonal anti-human neomycin phosphotransferase II (NPII) (Alpha Diagnostic International), rabbit polyclonal anti-Nluc antibody (Promega Corporaton), and polyclonal rabbit anti-*C. parvum* tryptophan synthase B (TrpB; B.S., unpublished observations) at 1:1,000, secondary antibodies were anti-mouse or anti-rabbit conjugated to Alexa488 or Alexa546 (Molecular Probes, Life Technologies) at a dilution of 1:1,000. DNA was visualized with DAPI ($2\,\mathrm{mg\,ml^{-1}}$). Images were collected on an Applied Precision Delta Vision inverted epifluorescence microscope at the UGA Biomedical Microscopy Core, deconvolved and adjusted for contrast using SoftWoRx software.

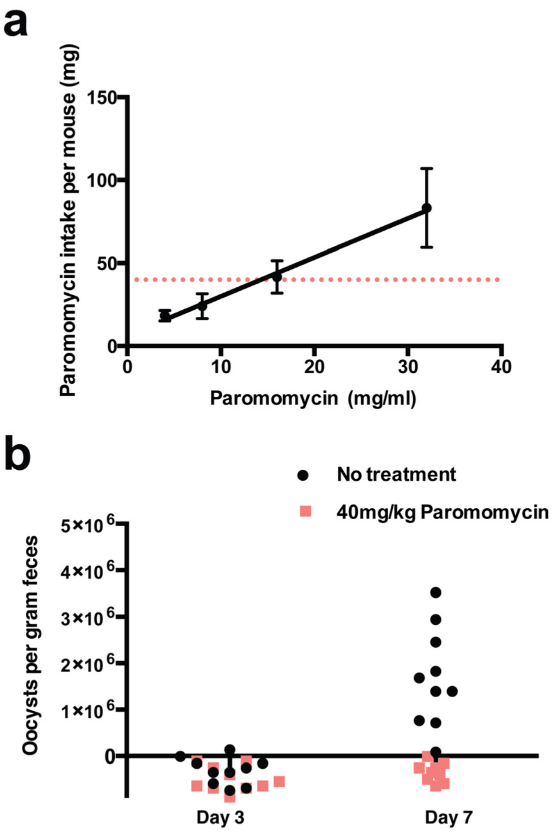
Statistical methods. All bar graphs depict the mean with standard deviations shown as error bars. Unless indicated otherwise, graphed data represent three technical replicates; each experiment was repeated at least twice and representative data are shown. No statistical tests were used to predetermine sample size. Unpaired *t*-tests were used appropriately to determine statistical significance and a P value <0.05 was considered significant. Assumptions for statistical tests were confirmed or corrected as described. No animals were excluded from experimental measurements.

31. Harb, O. S. & Roos, D. S. The Eukaryotic Pathogen Databases: a functional genomic resource integrating data from human and veterinary parasites. *Methods Mol. Biol.* **1201,** 1–18 (2015).
32. van Dooren, G. G., Tomova, C., Agrawal, S., Humbel, B. M. & Striepen, B. *Toxoplasma gondii* Tic20 is essential for apicoplast protein import. *Proc. Natl Acad. Sci. USA* **105,** 13574–13579 (2008).
33. Gubbels, M. J., Li, C. & Striepen, B. High-throughput growth assay for *Toxoplasma gondii* using yellow fluorescent protein. *Antimicrob. Agents Chemother.* **47,** 309–316 (2003).
34. Saeij, J. P., Boyle, J. P., Grigg, M. E., Arrizabalaga, G. & Boothroyd, J. C. Bioluminescence imaging of *Toxoplasma gondii* infection in living mice reveals dramatic differences between strains. *Infect. Immun.* **73,** 695–702 (2005).
35. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
36. Chakrabarti, K. *et al.* Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis. *RNA* **13,** 1923–1939 (2007).
37. Striepen, B. & Soldati, D. in *Toxoplasma gondii: The Model Apicomplexan — Perspective and Methods* (eds Weiss, L. M. & Kim, K.) 391–415 (Elsevier, 2007).
38. Sharling, L. *et al.* A screening pipeline for antiparasitic agents targeting cryptosporidium inosine monophosphate dehydrogenase. *PLoS Negl. Trop. Dis.* **4,** e794 (2010).
39. Bessoff, K., Sateriale, A., Lee, K. K. & Huston, C. D. Drug repurposing screen reveals FDA-approved inhibitors of human HMG-CoA reductase and isoprenoid synthesis that block *Cryptosporidium parvum* growth. *Antimicrob. Agents Chemother.* **57,** 1804–1814 (2013).
40. Mary, C. *et al.* Multicentric evaluation of a new real-time PCR assay for quantification of *Cryptosporidium* spp. and identification of *Cryptosporidium parvum* and *Cryptosporidium hominis*. *J. Clin. Microbiol.* **51,** 2556–2563 (2013).
41. Upton, S. J. in *Cryptosporidium and Cryptosporidiosis* (ed. Fayer, R.) 181–207 (CRC, 1997).

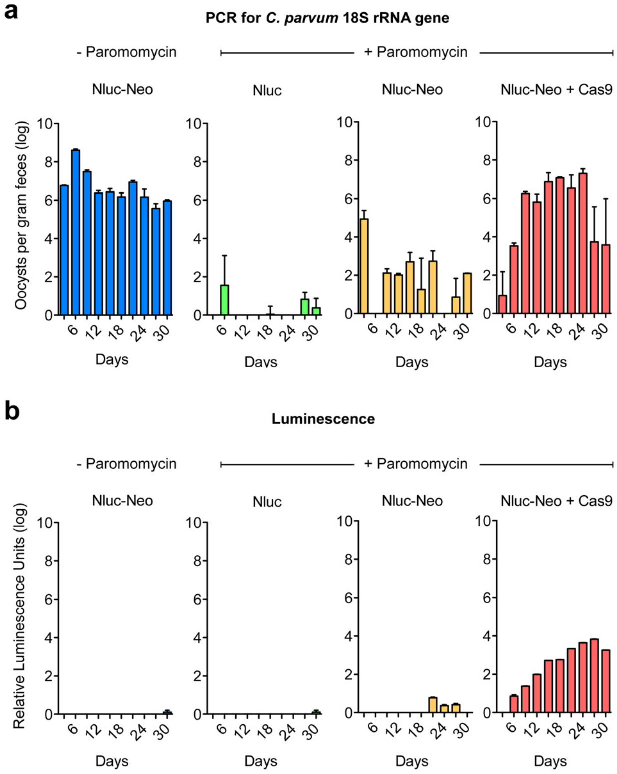


Extended Data Figure 1 | Optimization of sporozoite transfection. **a**, Ten-million sporozoites prepared in either cytomix (BTX) or Lonza Buffers SE, SF or SG (4D Nucleofection) were combined with 10 µg DNA (Eno_Nluc-GS-Nluc_Eno). Samples were electroporated using previously determined settings for BTX (1,500 V, 25 Ω, 25 µF) or various program settings for 4D Nucleofection as indicated. Parasites were added to cultures of HCT-8 cells and luciferase activity was read after 48 h. Bars represent average of two technical replicates. **b**, Transfection was further optimized by comparing the best preliminary settings (buffers SF and SG; programs EH 100 and EO 100) with additional pulse programs as indicated. Transfection was carried out as in **a**. Bars represent average of two technical replicates. **c**, Electroporation systems (BTX and 4D Nucleofection) were compared using the same number of *C. parvum* sporozoites and quantities of DNA using buffers and conditions optimized in **a** and **b**. Bars represent average of three technical replicates. Note about tenfold enhancement of transient transfection using 4D Nucleofection. The impact of electroporation on stable transformation cannot be assessed in this setup and may be higher. Experiments in **a** and **b** were done once for the purpose of optimization, while **c** was repeated three times; a single representative experiment is shown.

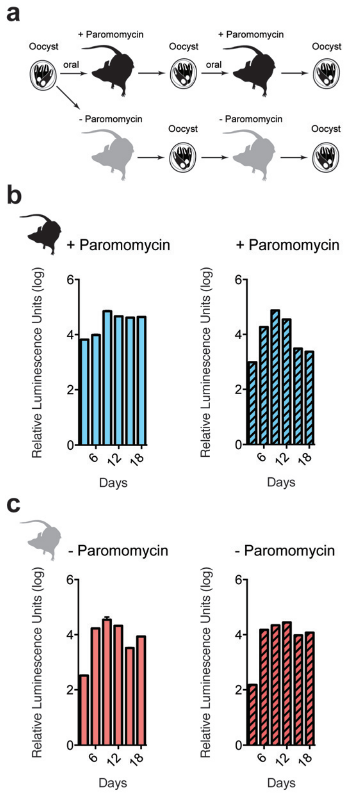
Extended Data Figure 2 | Direct surgical injection of transfected *C. parvum* sporozoites into the small intestine. Mice are shaved and anaesthetized with isofluorane (3% initially, then maintained at 1.5% for the surgery). The abdominal skin is disinfected with Betadine and a small incision is made into the peritoneum. Forceps are used to grasp the small intestine and 100 μl of PBS containing 10^7 transfected *C. parvum* sporozoites is injected into the lumen. The peritoneum and the abdominal skin are each sutured with 4-0 polydioxanone and mice are injected with meloxicam (1 mg kg^{-1}) subcutaneously. Each procedure takes around 15 min, and mice recover rapidly.



Extended Data Figure 3 | Optimization of paromomycin treatment of infected mice. **a**, Dosing of mice accounting for drug concentration, animal weight, and measured daily water consumption. At 16 mg ml^{-1} each mouse received 40 mg paromomycin daily (dotted line). **b**, This dose was found to be sufficient to decrease oocyst shedding in treated mice to background. By day 7 mice without paromomycin treatment shed large amounts of oocysts when compared to untreated mice. Treated mice showed no shedding above background. Oocysts were enumerated by high-throughput imaging assay. Five mice were analysed individually with two technical replicates.



Extended Data Figure 4 | Mouse model for selection of stable *C. parvum* transgenics. Repeat of the experiment described in Fig. 3b. **a**, Measurement of *C. parvum* infection using faecal PCR. **b**, Luminescence measurements. Note increasing luminescence from day 6 in parasites that received resistance and Cas9 plasmids. Mice were infected in groups of four per cage and pooled faeces was analysed for each cage (each measurement represents three technical replicates).



Extended Data Figure 5 | *C. parvum* maintains the stable transgene when passed serially in mice without paromomycin treatment. **a**, Mice were infected orally with 100,000 transgenic oocysts. **b**, **c**, Infected mice were then treated with paromomycin (**b**) or left untreated (**c**). Oocysts were purified from faecal collections by sucrose flotation and CsCl centrifugation, and used to infect a second cohort of mice. Again, each mouse received 100,000 transgenic oocysts and mice were treated or not. Faeces were tested for luminescence every 3 days. Each reading represents the pooled faecal sample from five mice with three technical replicates.

nature biotechnology

Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells

Van Trung Chu[1], Timm Weber[1], Benedikt Wefers[2,3], Wolfgang Wurst[2–4], Sandrine Sander[1], Klaus Rajewsky[1] & Ralf Kühn[1,2,5]

The insertion of precise genetic modifications by genome editing tools such as CRISPR-Cas9 is limited by the relatively low efficiency of homology-directed repair (HDR) compared with the higher efficiency of the nonhomologous end-joining (NHEJ) pathway. To enhance HDR, enabling the insertion of precise genetic modifications, we suppressed the NHEJ key molecules KU70, KU80 or DNA ligase IV by gene silencing, the ligase IV inhibitor SCR7 or the coexpression of adenovirus 4 E1B55K and E4orf6 proteins in a 'traffic light' and other reporter systems. Suppression of KU70 and DNA ligase IV promotes the efficiency of HDR 4–5-fold. When co-expressed with the Cas9 system, E1B55K and E4orf6 improved the efficiency of HDR up to eightfold and essentially abolished NHEJ activity in both human and mouse cell lines. Our findings provide useful tools to improve the frequency of precise gene modifications in mammalian cells.

The CRISPR-Cas9 systems (clustered, regularly interspaced, short palindromic repeats (CRISPR)–CRISPR-associated protein) represent a versatile tool for genome engineering, enabling the induction of site-specific genomic double-strand breaks (DSBs) by single guide RNAs (sgRNAs)[1]. In mammalian cells DSBs are mostly repaired by the nonhomologous end-joining (NHEJ) pathway[2,3], frequently leading to the loss of nucleotides from the ends of DSBs. This enables the efficient construction of knockout alleles through the induction of frameshift mutations[4]. By contrast, the alternative pathway of homology-directed repair (HDR) can be used for the introduction of precise genetic modifications such as codon replacements or reporter insertions by recombination with exogenous targeting vectors, serving as repair template[5]. We reasoned that the efficiency of HDR and thus the construction of precise genetic modifications could be boosted by the transient inhibition of NHEJ key molecules, similar to what has been observed for *Drosophila* embryos with a genetic DNA ligase IV deficiency[6].

To quantitatively determine the outcome of CRISPR-Cas-induced DSB repair, we first generated human HEK293 cells with a 'traffic light' reporter[7] (TLR) vector integrated into the adeno-associated virus integration site 1 (AAVS1) locus[8] (**Fig. 1a**). HEK293 cells were transfected with an AAVS1 targeting vector carrying the TLR insert and expression plasmids for Cas9 and an AAVS1-specific sgRNA (**Supplementary Fig. 1**). Upon selection and genotyping of transfected cells, we obtained heterozygous (AAVS1$^{TLR/+}$) and homozygous (AAVS1$^{TLR/TLR}$) targeted clones harboring the TLR construct in the AAVS1 locus (**Supplementary Fig. 2**). The reporter includes a CAG promoter for expression of a nonfunctional green fluorescent (Venus) gene, disrupted by the replacement of codons 117–152 with target sequences from the mouse *Rosa26* and *Rab38* locus, followed by coding regions for a self-cleaving 2A peptide and a red fluorescent (TagRFP) gene in a reading frame shifted by 2 bp (**Supplementary Table 1**). CRISPR-Cas9-induced DSBs in the target region that are repaired by means of NHEJ and cause deletions shift the translation to the frame of the 2A-TagRFP in about 1/3 of the mutagenic NHEJ events; this can be detected in reporter cells by the expression of RFP (**Fig. 1b**). If an intact Venus coding region is provided as repair template, cells that repair the DSBs by HDR express Venus. For activation of the TLR reporter, we designed two sgRNAs against the *Rosa26* target sequence, of which sgRosa26-1 showed a higher activity to induce deletions in the endogenous locus of mouse NIH3T3 cells (**Fig. 1c**).

Next, we transfected AAVS1$^{TLR/+}$ cells with an expression vector for Cas9, blue fluorescent protein (BFP) and sgRosa26-1, together with a linearized Venus donor plasmid. 72 h after transfection the cells were analyzed by FACS, gated for BFP$^+$ transfected cells, and the frequency of Venus$^+$ and of RFP$^+$ cells was determined. We observed 3% RFP$^+$ and 5% Venus$^+$ cells, indications of NHEJ or HDR repair events, respectively, as compared to 0.1% RFP$^+$ and 0.6% Venus$^+$ cells in a control lacking sgRosa26-1 (**Fig. 1d**). Of note, RFP$^+$ cells detected by the TLR assay represent only 1/3 of all mutagenic NHEJ events. Similar results were obtained with AAVS1$^{TLR/TLR}$ cells (**Supplementary Fig. 3**).

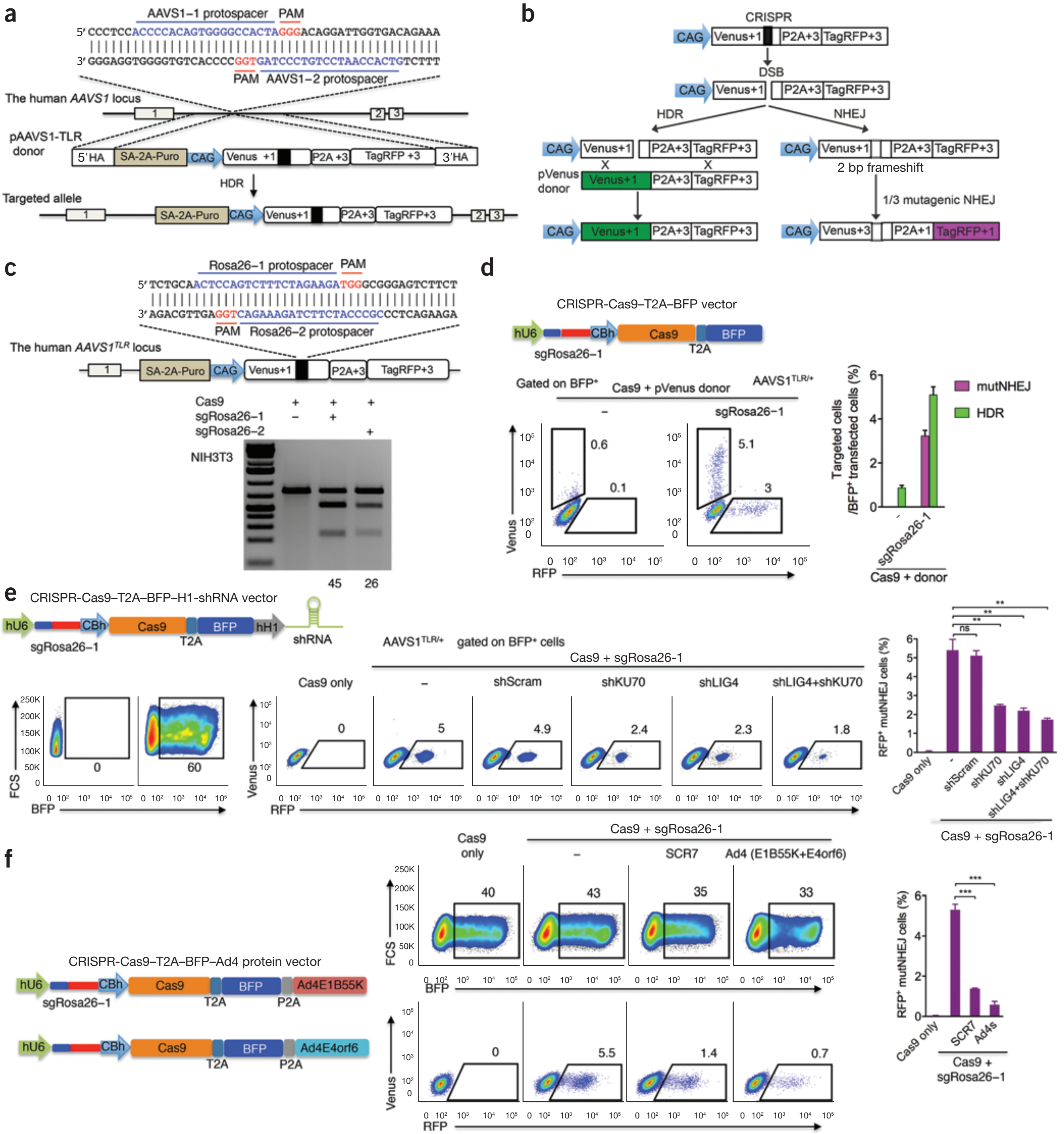
For suppression of key NHEJ pathway proteins[2,9] by short hairpin (sh) RNAs, we added a human H1 promoter to the sgRosa26-1/Cas9/BFP expression vector and inserted published shRNA sequences to knock down KU70, KU80 or DNA ligase IV. We first determined the extent of NHEJ suppression by transfection of AAVS1$^{TLR/+}$ cells with different

[1]Max-Delbrück-Center for Molecular Medicine, Berlin, Germany. [2]Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [3]Deutsches Zentrum für Neurodegenerative Erkrankungen e.V., Munich, Germany. [4]Munich Cluster for Systems Neurology (SyNergy), Ludwig-Maximilians-Universität München, Munich, Germany. [5]Berlin Institute of Health, Berlin, Germany. Correspondence should be addressed to R.K. (ralf.kuehn@mdc-berlin.de) or K.R. (klaus.rajewsky@mdc-berlin.de).

Received 13 February; accepted 18 March; published online 24 March 2015; doi:10.1038/nbt.3198

sgRosa26-1–Cas9-BFP-H1 shRNA expression vectors in the absence of a repair template. 72 h after transfection the samples were analyzed for RFP$^+$ cells and compared to controls that had either a scrambled shRNA or no shRNA. A substantial suppression of NHEJ repair was observed upon the individual or combined knockdown of KU70, KU80 or DNA ligase IV (**Fig. 1e** and **Supplementary Fig. 4**). Similar results were obtained in AAVS1$^{TLR/TLR}$ cell lines by the knockdown of KU70, KU80 or DNA ligase IV (**Supplementary Fig. 4**). The knockdown of ligase IV reduced its protein level in transfected AAVS1^{TLR} (BFP$^+$) cells by 70% (**Supplementary Fig. 5**).

As additional approaches to DNA ligase IV inhibition, we used the small-molecule inhibitor SCR7 (ref. 10) or the adenovirus 4 (Ad4) E1B55K and E4orf6 proteins, which mediate the ubiquitination and proteasomal degradation of DNA ligase IV[11,12]. For the coexpression of Ad4 proteins, sgRosa26-1, Cas9 and BFP from a vector pair, we linked the Ad4 E1B55K or the Ad4 E4orf6 gene by self-cleaving 2A peptide sequences to BFP (**Fig. 1f**). Mono- or biallelic AAVS1^{TLR} cell lines were transfected with either the sgRosa26-1–Cas9-BFP expression vector (in the presence or absence of SCR7 inhibitor) or both sgRosa26-1/Cas9/BFP/Ad4 E1B55K and E4orf6 plasmids. The presence of SCR7



reduced the fraction of RFP$^+$ cells in the BFP$^+$ population fourfold whereas the coexpression of Ad4 proteins led to an eightfold reduction, compared to controls lacking inhibitor or Ad4 proteins (**Fig. 1f** and **Supplementary Fig. 6**). The coexpression of Ad4 proteins reduced the level of DNA ligase IV protein in transfected AAVS1^{TLR} (BFP$^+$) cells by 93% (**Supplementary Fig. 5**). Overall, these results show that the NHEJ repair of CRISPR-Cas9-induced DSBs can be suppressed by targeting DNA ligase IV using RNA interference, SCR7 or, most efficiently, the Ad4 proteins.

To assess the effect of NHEJ suppression on HDR, we transfected AAVS1$^{TLR/+}$ cells with Venus repair template together with sgRosa26-1/Cas9/BFP vector (with and without SCR7) or with sgRosa26-1/Cas9/BFP vectors including either the Ad4 proteins or shRNA constructs targeting KU70, KU80 or DNA ligase IV. After 72 h the frequency of RFP$^+$ and Venus$^+$ cells within BFP$^+$ cells was analyzed by FACS (**Fig. 2a**,**b** and **Supplementary Fig. 7**). Venus$^+$ (HDR) cells increased from 5% for sgRosa26-1/Cas9/BFP alone, to 8–14% in the presence of single shRNAs against KU70, KU80 or DNA ligase IV, to 25% in the presence of shRNAs against KU70 and DNA ligase IV or 1 μM of the inhibitor SCR7, and further to 36% upon the coexpression of the Ad4 proteins. Thus, HDR efficiency was enhanced up to fivefold in the presence of KU70 and ligase IV shRNAs or SCR7, and up to sevenfold by the Ad4 protein pair (**Fig. 2c** and **Supplementary Fig. 8**).

Titration of SCR7 on AAVS1$^{TLR/+}$ cells showed an optimal effect at 1 μM concentration (**Supplementary Fig. 9**). For cells in the presence of two shRNAs, SCR7 or Ad4 proteins, we noticed diminished fluorescence signals within the population of Venus$^+$ cells at 72 h after transfection (**Fig. 2a** and **Supplementary Figs. 7**,**9c**), indicating reduced Venus expression in cells undergoing NHEJ blockade, possibly caused by local chromatin remodeling through an extended DNA damage response[13,14]. However, Venus expression was normal in clones established from AAVS1$^{TLR/+}$ cells targeted in the presence of Ad4 proteins, indicating that this effect is only transient (**Supplementary Fig. 9d**). From the sample expressing the Ad4 proteins, Venus$^+$ cells were sorted, and we established 24 clones to confirm the integrity of the repaired TLR loci using PCR and sequence analysis (**Supplementary Table 2**). In contrast to the increase of Venus$^+$ cells, RFP$^+$ cells decreased from 3% in the controls to 1.7%, 1.4% or 0.6% in the presence of shRNAs, SCR7 or Ad4 proteins, respectively (**Fig. 2a**,**b**). Whether the residual NHEJ activity relies on the KU- and ligase IV–independent alternative end-joining mechanism[15] remains to be determined.

Assessing the influence of the lengths of homology regions of the repair template on HDR efficiency, we generated donor templates with 3′ homology regions shortened from 1,450 bp as in the original donor template to 350 bp (**Fig. 2d** and **Supplementary Fig. 10**). We transfected AAVS1$^{TLR/+}$ cells with sgRosa26-1/Cas9/BFP expression vector and each of the various donor templates, with or without the coexpression of Ad4 proteins. FACS analysis revealed a reduced targeting frequency (2%) for the donor with a 350 bp 3′ homology region whereas the other molecules showed HDR efficiencies in the range of 5% (**Fig. 2e**). In the presence of Ad4 proteins the frequency of Venus$^+$ cells increased robustly up to 25% for the 350-bp donor and to 30% for the other donor molecules. In line with the previous results, the frequency of RFP$^+$ cells was strongly reduced by the coexpression of Ad4 proteins (**Fig. 2f**). Thus, PCR-generated fragments with combined homology regions of >1 kb are effective donors for HDR (**Fig. 2e**,**f** and **Supplementary Fig. 10**).

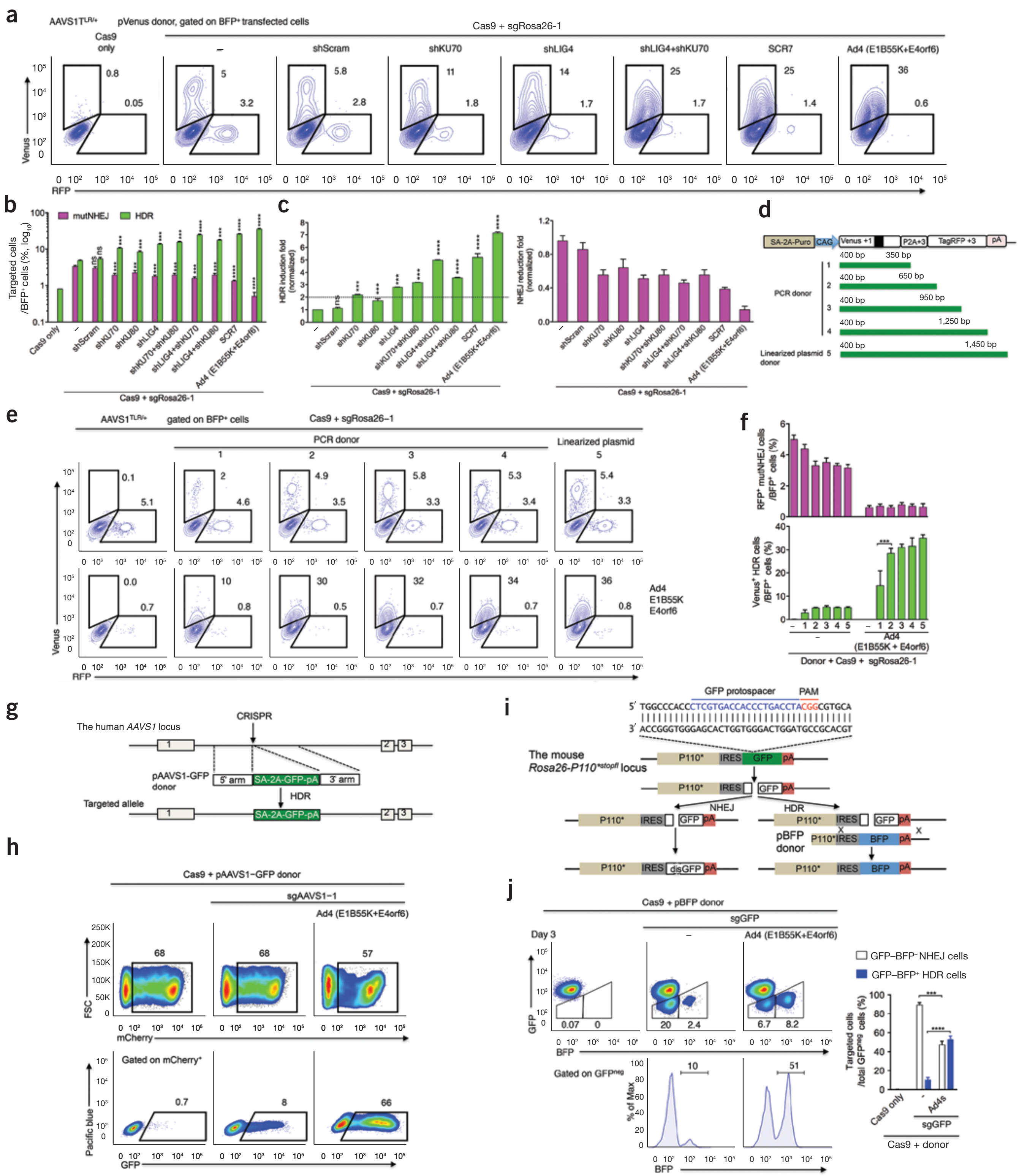
Applying our approach to an endogenous genomic locus, we inserted a GFP reporter gene into the *AAVS1* locus of HEK293 cells. The AAVS1-SA-T2A-GFP targeting vector includes AAVS1 homology regions flanking a splice acceptor site and a 2A peptide sequence linked to GFP, enabling reporter expression by the AAVS1-derived transcript (**Fig. 2g**). HEK293 cells were cotransfected with an AAVS1-specific sgRNA/Cas9/mCherry (AAVS1-1/Cas9/mCherry) expression plasmid and the AAVS1-SA-T2A-GFP targeting vector, with or without coexpression of Ad4 proteins. 72 h after transfection the mCherry$^+$ transfected population exhibited 8% GFP$^+$ cells upon expression of sgAAVS1-1/Cas9 alone whereas the coexpression of Ad4 proteins raised the frequency of the cells to 66% (**Fig. 2h**). Thus, in line with the TLR results, we observed an eightfold stimulation of gene targeting at the *AAVS1* locus by the coexpression of Ad4 proteins. FACS sorting and cloning GFP$^+$ cells 48 h after transfection, we found correct GFP gene integration in 45 of 48 samples (95%), but only in 60% of the clones derived from the sample without Ad4 proteins (**Supplementary Fig. 11** and **Supplementary Table 3**), the remaining cells presumably representing random integrants.

Coexpression of Ad4 proteins also promoted HDR efficiency in a mouse Burkitt lymphoma (BL) cell line containing an activated PI3 kinase α-subunit[16] linked to an IRES element and a GFP reporter in the *Rosa26* locus. We targeted the GFP using a specific sgRNA and a promoterless donor vector replacing GFP by BFP (**Fig. 2i**). The fraction of BFP$^+$ cells indicates HDR efficiency whereas the fraction of GFP$^-$ cells indicates

Figure 1 Insertion of a traffic light reporter into the *AAVS1* locus of HEK293 cells and suppression of the NHEJ pathway. (**a**) Strategy for insertion of the TLR construct into the *AAVS1* locus using CRISPR-Cas9 in human HEK293 cells. In the targeted sequence, the AAVS1-specific sgRNA is indicated in blue and the PAM signal is shown in red. The pAAVS1-TLR targeting vector includes homology arms (HA) of 800 bp flanking a splice acceptor (SA)-2A-puromycin element and the traffic light reporter insert comprising a CAG promoter and a Venus gene inactivated by the replacement of 36 codons with target sequences from the mouse Rosa26/Rab38 loci (black insert). (**b**) Diagram of the TLR system. CRISPR-Cas9-induced DSBs in the target region, repaired by NHEJ resulting in deletions that shift translation by 2 bp, led to RFP expression. An intact Venus coding region serves as repair template for HDR, leading to Venus expression. (**c**) Strategy to target the human *AAVS1^{TLR}* locus. Two different protospacers targeting the mouse *Rosa26-derived* sequence that interrupts the Venus gene are indicated in blue, with PAM signals in red. NIH3T3 cells were transfected by empty vectors or vectors expressing sgRosa26-1 or sgRosa26-2/Cas9-T2A-mCherry to test cutting efficiency in the endogenous *Rosa26* locus. 48 h after transfection, mCherry$^+$ cells were sorted and PCR and T7EI assays performed. The percentage of indels was quantified by the ImageJ software. (**d**) AAVS1^{TLR} cells were co-transfected with linearized Venus repair vector and Cas9-BFP expression plasmids with or without (–) sgRosa26-1. 72 h after transfection, flow cytometric analysis of BFP$^+$ gated cells displayed 5% of Venus$^+$ (HDR repair) cells and 3% of RFP$^+$ (NHEJ repair) cells. The graphs represent triplicate data from one of three independent experiments with similar results, shown as mean ± s.d. (**e**) Inhibition of the NHEJ pathway using gene silencing. Scheme of the sgRosa26-1/Cas9-2A-BFP-H1shRNA expression vector. AAVS1^{TLR} cells were transfected with sgRosa26-1/Cas9-2A-BFP (-) or sgRosa26-1/Cas9-2A-BFP-H1-shScrambled, -shKU70, -shLIG4 or -shKU70 or -shLIG4. 3 days later, these cells were analyzed by flow cytometry, gated on BFP$^+$ transfected cells, and the percentage of RFP$^+$ cells was determined. The graphs represent triplicate data from 1 of 3 independent experiments with similar results, shown as mean ± s.d. (**f**) Suppression of NHEJ repair using the ligase IV inhibitor SCR7 or the coexpression of ligase IV degrading Ad4 E1B55K and E4orf6 (Ad4s) proteins. AAVS1^{TLR} cells were transfected with sgRosa26-1/Cas9-2A-BFP expression vector (-) or transfected and cultured with SCR7 or were transfected with the sgRosa26-1/Cas9-2A-BFP-Ad4-E1B55K or -E4orf6 expression vectors. The samples were analyzed 3 days later by flow cytometry, gating on the BFP$^+$ transfected cells, and the percentage of RFP$^+$ cells was determined. The graphs represent triplicate data from one of six independent experiments with similar results, shown as mean ± s.d. Significance was calculated using the Student's *t*-test: $^{**}P < 0.01$, $^{***}P < 0.001$, ns, not significant.

NHEJ-mediated deletion events. Mouse BL cells were electroporated with Cas9 and the BFP replacement vector alone or with GFP-specific sgRNA/Cas9 expression plasmid with or without the coexpression of Ad4 proteins. Electroporation with sgGFP and Cas9 led to a transfection efficiency of 40%, as determined by mCherry expression after 24 h. About half of the transfected cells (22% in total) lost GFP expression after 72 h (**Supplementary Fig. 12**) and 10% of the GFP⁻ cells were BFP⁺ (**Fig. 2j**). The addition of Ad4 proteins reduced transfection efficiency to 27%, but again about half of the transfected cells (14% in total) lost GFP expression after 72 h (**Supplementary Fig. 12**). Notably, 50% of the GFP⁻ cells were now BFP⁺ (**Fig. 2j**), indicating a fivefold stimulation of HDR by Ad4 proteins.

To assess whether a point mutation can be corrected in a tumor cell line, we reverted the T24P codon replacement in the *Foxo1* gene



of mouse BL cells, which renders FOXO1 resistant to AKT-dependent phosphorylation. We transfected BL cells with a mutation-specific sgRNA and a second sgRNA recognizing the first intron, expression vectors for Cas9 and fluorescent reporters without or together with Ad4 proteins and a targeting vector containing the reverted codon 24 (and harboring a silent nucleotide replacement) and a puromycin-resistance gene (**Supplementary Fig. 13**). At day 2 transfected cells were isolated by FACS sorting, subjected to puromycin selection, and 53 or 39 clones derived from single cells were established from the sample without or with coexpression of Ad4 proteins, respectively. Genotyping of these clones by PCR and sequence analysis showed that 43 of the 53 clones derived without Ad4 proteins were mutants, 33 of these were heterozygous and 10 clones (19%) were targeted on both alleles. In the presence of Ad4 proteins, we found that all 39 clones were targeted, 24 of these were heterozygous and 15 clones (38%) homozygous mutants. Thus, despite the expected high targeting rate achieved with a selectable donor vector, the coexpression of Ad4 proteins further increased the targeting efficiency to 100% and doubled the net yield of homozygous, targeted clones.

Figure 2 Enhancement of HDR for CRISPR-Cas9-induced precise gene targeting. (**a**) Improvement of HDR efficiency by suppression of NHEJ key molecules. AAVS1^{TLR} cells were cotransfected with linearized Venus repair vector and sgRosa26-1/Cas9/BFP expression plasmid together with shRNA cassettes, SCR7 inhibitor or the Ad4 E1B55K/E4orf6 proteins. The frequency of RFP$^+$ and Venus$^+$ cells within the transfected BFP$^+$ population was determined by flow cytometry. The data represent one of four independent experiments with similar results. (**b**) The graph summarizes the frequency of RFP$^+$ (pink bars, significance compared to the (–) sample) and of Venus$^+$ (green bars; significance compared to the (–) sample) cells determined as in **a**. The *y* axis is represented as $\log_{10}$ scale. The bars represent mean values ± s.d. (**c**) Relative increase of HDR efficiency (significance compared to the (–) sample) and of NHEJ suppression normalized to the control transfected with sgRosa26-1/Cas9 and targeting vector alone (–). The dotted line indicates a more than twofold increase of HDR. The bars represent mean values ± s.d. (**d**) Use of PCR-generated donor templates. Scheme of the TLR construct and of PCR donor fragments, having a constant 5′ homology region of 400 bp whereas the length of the 3′ homology region varies from 350 bp to 1,250 bp; as control the linearized promoterless Venus repair vector was used. (**e**) AAVS1^{TLR} cells were cotransfected with sgRosa26-1/Cas9/BFP expression vector, the same molar amounts (800 fmol) of PCR donors or linearized Venus repair vector without or together with Ad4 E1B55K/E4orf6 proteins. The percentage of Venus$^+$ and RFP$^+$ cells was analyzed by flow cytometry 3 days after transfection. (**f**) The graph summarizes the frequency of RFP$^+$ (purple bars) and Venus$^+$ (green bars) cells determined as in **e**. The data represent one of two independent experiments with similar results. The bars represent mean values ± s.d. (**g**) Strategy for insertion of a GFP reporter gene into the human *AAVS1* locus using CRISPR-Cas9 in human HEK293 cells. The CRISPR-Cas9-targeted site is shown in **Figure 1a**; in the AAVS1-GFP targeting vector the GFP gene is flanked by AAVS1 homology regions of 800 bp. (**h**) HEK293 cells were cotransfected with linearized AAVS1-GFP targeting vector and with sgAAVS1-1/Cas9/mCherry expression vector or with sgAAVS1-1/Cas9/mCherry-Ad4-E1B55K and -E4orf6 expression vectors. At day 3, the frequency of GFP$^+$ cells within the population of mCherry+ transfected cells was analyzed using the flow cytometric data; shown is one of triplicate samples obtained from one of two independent experiments. (**i**) Fluorescent reporter replacement in the mouse BL cell line. The cell line harbors an activated PI3 kinase (P110*)-IRES-GFP-pA cassette in the *Rosa26* locus. In the targeted GFP sequence, the sgGFP target sequence is highlighted in blue and the PAM element in red. DSBs repaired by the NHEJ pathway led to the inactivation of GFP. DSBs repaired by HDR with the pBFP donor template led to the replacement of GFP by the BFP reporter gene. (**j**) Mouse BL cells were electroporated with Cas9/mCherry or sgGFP/Cas9/mCherry expression vector and the pBFP donor plasmid without or together with coexpression of the Ad4 E1B55K/E4orf6 proteins. The frequency of GFP$^-$BFP$^-$ (white bars) and GFP$^-$BFP$^+$ (blue bars) cells was analyzed at day 3. The graph summarizes triplicate results from one of three independent experiments with similar results. The bars represent mean values ± s.d. Significance was calculated using the Student's *t*-test: $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, ns, not significant.

In summary, we show that for CRISPR-Cas9-induced mutagenesis the suppression of the NHEJ key enzyme DNA ligase IV is an effective way for engineering precisely targeted mutations into the genome of mammalian cells. The activity of ligase IV can be blocked by gene silencing, small-molecule inhibition or proteolytic degradation, offering diverse approaches for the optimal delivery into target cells. For the proteolytic degradation of DNA ligase IV, we selected the E1B55K and E4orf6 proteins of adenovirus 4, shown to exert minimal influence on other cellular substrates such as Mre11 or p53, which are co-targeted by many other serotypes[17]. Nevertheless, we cannot exclude the possibility that HDR stimulation by the Ad4 proteins is mediated by the combined suppression of DNA ligase IV and other regulatory proteins. It could be of further interest to compare the effect on HDR of the E1B55K and E4orf6 proteins of additional adenoviral serotypes and species identified from humans[18] and other vertebrates[19]. Furthermore, the use of Ad4 proteins may be also beneficial for the construction of targeted mouse mutants, as recently shown for zygotes cultured in the presence of SCR7 (ref. 20). In populations of cells transfected with a sgRNA, Cas9 and Ad4 protein expression plasmids, we presently reach knock-in frequencies of 50–66%. By delivering Cas9 and sgRNAs as synthetic RNAs it may be possible, as shown for human induced pluripotent stemcells[21], to further enhance gene targeting efficiencies. It will be interesting to apply CRISPR-Cas9 mutagenesis combined with NHEJ suppression also to early embryos of other model organisms and to primary mammalian cells to achieve gene corrections.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank K. Petsch and H.P. Rahn from the FACS core facility for excellent technical support. This work was supported by the European Research Council (ERC advanced grant ERC-AG-LS6, to K.R.), the German Ministry of Education and Research within the VIP program (TAL-CUT 03V0261, to W.W. and R.K.) and the European Union within the EUCOMMTOOLS project (FP7-HEALTH-F4-2010-261492, to W.W.). The plasmids MSCV-IRES-mCherry and MSCV-IRES-BFP were kind gifts of F. Rosenbauer and M. Janz (Charite, Berlin). We are grateful to H. Ploegh for communicating to us unpublished results and support in coordinating publication.

AUTHOR CONTRIBUTIONS

V.T.C., K.R. and R.K. conceived and designed the project, V.T.C. and T.W. acquired the data, V.T.C., K.R., T.W. and R.K. analyzed and interpreted the data, B.W., S.S. and W.W. provided materials, and V.T.C., K.R. and R.K. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
2. Lieber, M.R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* **79**, 181–211 (2010).
3. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst.)* **7**, 1765–1771 (2008).
4. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
5. San Filippo, J., Sung, P. & Klein, H. Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem.* **77**, 229–257 (2008).

6. Bozas, A., Beumer, K.J., Trautman, J.K. & Carroll, D. Genetic analysis of zinc-finger nuclease-induced gene targeting in *Drosophila*. *Genetics* **182**, 641–651 (2009).
7. Certo, M.T. *et al*. Tracking genome engineering outcome at individual DNA breakpoints. *Nat. Methods* **8**, 671–676 (2011).
8. Samulski, R.J. *et al*. Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *EMBO J.* **10**, 3941–3950 (1991).
9. Panier, S. & Durocher, D. Push back to respond better: regulatory inhibition of the DNA double-strand break response. *Nat. Rev. Mol. Cell Biol.* **14**, 661–672 (2013).
10. Srivastava, M. *et al*. An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell* **151**, 1474–1487 (2012).
11. Cheng, C.Y. *et al.* The E4orf6/E1B55K E3 ubiquitin ligase complexes of human adenoviruses exhibit heterogeneity in composition and substrate specificity. *J. Virol.* **85**, 765–775 (2011).
12. Forrester, N.A. *et al*. Serotype-specific inactivation of the cellular DNA damage response during adenovirus infection. *J. Virol.* **85**, 2201–2211 (2011).
13. Adam, S. & Polo, S.E. Blurring the line between the DNA damage response and transcription: the importance of chromatin dynamics. *Exp. Cell Res.* **329**, 148–153 (2014).
14. Zhu, Q. & Wani, A.A. Histone modifications: crucial elements for damage response and chromatin restoration. *J. Cell. Physiol.* **223**, 283–288 (2010).
15. Frit, P., Barboule, N., Yuan, Y., Gomez, D. & Calsou, P. Alternative end-joining pathway(s): bricolage at DNA breaks. *DNA Repair (Amst.)* **17**, 81–97 (2014).
16. Srinivasan, L. *et al.* PI3 kinase signals BCR-dependent mature B cell survival. *Cell* **139**, 573–586 (2009).
17. Cheng, C.Y. *et al.* The E4orf6/E1B55K E3 ubiquitin ligase complexes of human adenoviruses exhibit heterogeneity in composition and substrate specificity. *J. Virol.* **85**, 765–775 (2011).
18. Ghebremedhin, B. Human adenovirus: viral pathogen with increasing importance. *Eur. J. Microbiol. Immunol.* **4**, 26–33 (2014).
19. Benkő, M. et al. in Virus Taxonomy, Eighth Report of the International Committee on Taxonomy of Viruses (Elsevier Academic Press, 2005).
20. Singh, P., Schimenti, J.C. & Bolcun-Filas, E. A mouse geneticist's practical guide to CRISPR applications. *Genetics* **199**, 1–15 (2015).
21. González, F. *et al.* An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* **15**, 215–226 (2014).

ONLINE METHODS

Traffic light reporter construct. The traffic light reporter (TLR) expression construct was assembled by cloning of PCR fragments encoding a defective Venus (codons 117–152 replaced by a 52-bp segment derived from the mouse *Rosa26* locus (sgRNA target sequence underlined) and a 56-bp segment from the mouse *Rab38* gene) linked to the 2A peptide from *Thosea asigna*[22] and the coding region of TagRFP[23] in a 2-bp shifted reading frame (+3), cloned in between the CAG promoter and the polyA region of the bovine hGH gene. The CAG promoter was deleted from this plasmid to derive the traffic light targeting vector (**Supplementary Table 1**).

Cell culture and reagents. Wild-type, AAVS1^{TLR} HEK293 and mouse NIH3T3 cells were maintained in DMEM (Gibco) supplied with 15% FBS (Gibco), cells were passaged three times per week. The mouse Burkitt lymphoma cell line, generated from a Burkitt-like mouse lymphoma[16] was maintained in DMEM supplied with 15% FBS, 2 mM HEPES (Gibco), 2 mM sodium pyruvate (Gibco), 2 mM L-glutamine (Gibco), and 1× NAA (Gibco), beta-mercaptoethanol (Sigma) and passaged four times per week. For puromycin selection, mCherry^{+} cells were sorted, seeded at 10^3 cells/well and selected with 3 μg/ml of Puromycin for 2 weeks. Then colonies were counted and single cells were sorted. The SCR7 inhibitor was purchased (Xcess Biosciences, San Diego, USA), 12 h after transfection these cells were maintained in complete medium supplied with 1 μM SCR7 inhibitor until analysis. At SCR7 concentrations of 60 μM and 10 μM, we observed a reduction of transfection efficiency and of cell viability.

Donor vectors and CRISPR-Cas9-T2A-reporter vectors. To generate the CRISPR-Cas9-T2A reporter vector, we amplified T2A-mCherry and T2A-BFP fragment by overlapping PCR and cloned them into FseI/EcoRI sites of plasmid pX330 (Addgene, #42230). The mCherry and BFP templates were derived from the plasmids MSCV-IRES-mCherry and MSCV-IRES-BFP, respectively, a kind gift of Frank Rosenbauer and Martin Janz (Charite, Berlin). To generate the AAVS1-SA-2A-Puro-TLR targeting vector, a CAG-Venus+1-P2A+3-mtagRFP+3 cassette was cloned into SalI/NotI sites of AAVS1-SA-2A-Puro targeting vector (Addgene, #22075). In addition, a SA-2A-GFP fragment was generated by overlapping PCR and inserted into the XhoI/SalI sites of the AAVS1 targeting vector, with the GFP template derived from a pRosa26-IRES-GFP plasmid (V.T.C., unpublished). For reporter replacement in the mouse cell line, the P110*-IRES-BFP-pA construct was generated by overlapping PCR and cloned into the pSTblue-1 sequencing plasmid (Novagen). For PCR donor templates, fragments were amplified with Herculase II Fusion DNA Polymerase (Agilent Technology) from pVenus+1-P2A+3-tagRFP plasmid using the same forward primer and different reverse primers (**Supplementary Table 4**).

Generation of CRISPR-Cas9 vector expressing shRNA or Ad4 proteins. To generate CRISPR-Cas9-T2A-BFP/mCherry-hH1-shRNA vectors, we generated the human H1 promoter and MCS sequence by overlapping PCR and cloned it into the NotI site of the CRISPR-Cas9-T2A-BFP/mCherry plasmid. To obtain CRISPR-Cas9-T2A-BFP/mCherry-P2A-Ad E1B55K or E4orf6 plasmids, the coding regions for adenoviral serotype 4 proteins were synthesized as mammalian codon–optimized sequences (**Supplementary Table 1**) by Genscript (Piscataway, NJ, USA). Using these genes and BFP and mCherry template plasmids, we amplified BFP/mCherry-P2A-Ad4 E1B55K and BFP/mCherry-P2A-Ad4 E4orf6 fragments and cloned them into the NheI/EcoRI sites of the CRISPR-Cas9-T2A-reporter plasmid by Gibson assembly (New England Biolabs, E2611S).

sgRNA and shRNAs. sgRNAs were designed based on unique sequences with 20 nt and as the last nucleotide before the PAM signal an A or G was selected. The target sequences should hybridize with the sgRNA scaffold only at low energy as predicted by the Mfold web server (http://mfold.rna.albany.edu/?q=mfold/rna-folding-form). Complementary oligonucleotides were ordered separately, annealed, phosphorylated and cloned into the BbsI sites of the CRISPR-Cas9-T2A-reporter plasmid (**Supplementary Table 4**). For shRNA silencing, shRNA-targeted sequences were selected from previous reports[24–29] (**Supplementary Table 4**). Complementary oligonucleotides were ordered separately, annealed, phosphorylated and cloned into the BamHI/AflII sites of the CRISPR-Cas9-T2A-reporter-hH1 plasmid.

Transfection and electroporation. Human HEK293 and mouse NHI3T3 cells were plated into 24-well or 6-well plates at 1 day before transfection. On the day of transfection, these cells were supplied with new complete medium and the DNA mixed with FuGENE HD Reagent (Promega) in Opti-MEM (Invitrogen) according to the manufacturer's introduction. After 15 min of incubation at room temperature, the mixture was dropped slowly into the well. For electroporation, mouse BL cells were harvested and counted, $1–2 \times 10^6$ cells resuspended with 3 μg plasmid DNA in 100 ml electroporation buffer and transferred to a 0.2 cm cuvette (Sigma) and electroporated using a Nucleofactor device (Lonza). Then, cells were transferred into prewarmed complete medium.

Cell sorting and flow cytometry. For single-cell cloning, single cells were sorted into 96-well plates with 150 ml complete medium supplied with 10 μg/ml Gentamycin (Lonza). These plates were briefly centrifuged and incubated at 37 °C, 5% CO_2, the single-cell clones were evaluated 3 days after sorting to exclude multiple cell contamination. Cells were cultured until confluence and duplicated for genotyping PCR. For the bulk sorting, the reporter-positive cells were sorted into 15-ml Falcon tubes with complete medium, cells were centrifuged and further cultured or used for the isolation of genomic DNA. For flow cytometry analysis, HEK293 cells were trypsinized and resuspended in PBS/1% BSA FACS buffer and analyzed with a Fortessa machine (Becton Dickinson). Mouse BL cells were harvested, centrifuged and resuspended in PBS/1% BSA FACS buffer.

Genomic DNA isolation, PCR and T7EI assay. Reporter^{+} cells were cultured and harvested at different time points. Single-cell clones were duplicated in 96-well plates. Genomic DNA was extracted using the QuickExtract DNA extraction kit (Epicentre) following the manufacturer's instruction. For T7EI assay, PCR was done using Herculase II Fusion DNA Polymerase (Agilent Technology) with PCR gene-specific primers (**Supplementary Table 4**) using the following conditions: 98 °C for 3 min; 35–37 cycles (95 °C for 20 s, 60 °C for 20 s, 72 °C for 20 s) and 72 °C for 3 min. PCR products were run on 2% agarose gels, purified, denatured, annealed and treated with T7EI (New England Biolabs). Cleaved DNA fragments were separated on 2% agarose gels and the DNA concentration of each band was quantified using the ImageJ software. Percent values of indels were calculated as described[30]. For genotyping PCR, genomic DNA was amplified using DreamTaq DNA Polymerase (Thermo Scientific) with primers listed in **Supplementary Table 4**.

DNA sequencing. PCR products were directly sequenced by specific primers or cloned into the pSTBlue-1 Blunt vector (Novagen) following the manufacturer's protocol. Plasmid DNAs were isolated using the NucleoSpin Plasmid (Macherey-Nagel). Plasmids were sequenced using T7 forward primer (5′-TAATACGACTCACTATAGGG-3′) by the Sanger method (LGCgenomics, Berlin, Germany).

Western blot analysis. Transfected AAVS1^{TLR} Reporter (BFP^{+}) cells were isolated by FACS and 10×10^6 cells were lysed on ice in RIPA buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl 1, mM EDTA, 1% NP-40, 0.1% SDS, 0.1% sodium deoxycholate) for 20–30 min in the presence of protease inhibitors (Roche). The whole-cell lysates were centrifuged for 10 min at 14,000 r.p.m. The supernatants were transferred into new tubes and protein concentrations were determined using the BCA protein assay (Bio-Rad). The lysates were boiled at 100 °C for 5 min and loaded on SDS-PAGE gels. Blots were probed with anti-DNA ligase IV (H-300, Santa Cruz Biotechnology) and anti-beta-actin (AC-74, Sigma) antibodies. Blots were developed with secondary goat anti-rabbit IgG HRP (Southern Biotech) or anti-mouse IgG HRP (Southern Biotech) and bands visualized using the ECL detection kit (GE Healthcare).

22. Donnelly, M.L. *et al.* The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. *J. Gen. Virol.* **82**, 1027–1041 (2001).
23. Merzlyak, E.M. *et al.* Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat. Methods* **4**, 555–557 (2007).
24. Martinez, J.J., Seveau, S., Veiga, E., Matsuyama, S. & Cossart, P. Ku70, a component of DNA-dependent protein kinase, is a mammalian receptor for Rickettsia conorii. *Cell* **123**, 1013–1023 (2005).

doi:10.1038/nbt.3198

25. Com, E. *et al.* Nerve growth factor receptor TrkA signaling in breast cancer cells involves Ku70 to prevent apoptosis. *Mol. Cell. Proteomics* **6**, 1842–1854 (2007).
26. Li, B., Reddy, S. & Comai, L. Depletion of Ku70/80 reduces the levels of extrachromosomal telomeric circles and inhibits proliferation of ALT cells. *Aging* **3**, 395–406 (2011).
27. Munakata, Y. *et al.* Ku80 autoantigen as a cellular coreceptor for human parvovirus B19 infection. *Blood* **106**, 3449–3456 (2005).
28. Muylaert, I. & Elias, P. Knockdown of DNA ligase IV/XRCC4 by RNA interference inhibits herpes simplex virus type I DNA replication. *J. Biol. Chem.* **282**, 10865–10872 (2007).
29. Windhofer, F., Wu, W. & Iliakis, G. Low levels of DNA ligases III and IV sufficient for effective NHEJ. *J. Cell. Physiol.* **213**, 475–483 (2007).
30. Cong, L. et al. Multiplex genome engineering using CRISPR-Cas systems. *Science* **339**, 819–823 (2013).

R.K. and W.W. are inventors on a patent application entitled "Improved recombination efficiency by inhibition of NHEJ DNA repair" (EP2718446).

doi:10.1038/nature11927

A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity

Kimberley D. Seed[1], David W. Lazinski[1], Stephen B. Calderwood[2,3] & Andrew Camilli[1]

Bacteriophages (or phages) are the most abundant biological entities on earth, and are estimated to outnumber their bacterial prey by tenfold[1]. The constant threat of phage predation has led to the evolution of a broad range of bacterial immunity mechanisms that in turn result in the evolution of diverse phage immune evasion strategies, leading to a dynamic co-evolutionary arms race[2,3]. Although bacterial innate immune mechanisms against phage abound, the only documented bacterial adaptive immune system is the CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated proteins) system, which provides sequence-specific protection from invading nucleic acids, including phage[4–11]. Here we show a remarkable turn of events, in which a phage-encoded CRISPR/Cas system is used to counteract a phage inhibitory chromosomal island of the bacterial host. A successful lytic infection by the phage is dependent on sequence identity between CRISPR spacers and the target chromosomal island. In the absence of such targeting, the phage-encoded CRISPR/Cas system can acquire new spacers to evolve rapidly and ensure effective targeting of the chromosomal island to restore phage replication.

Vibrio cholerae serogroup O1 is the primary causative agent of the severe diarrhoeal disease cholera, and lytic *V. cholerae* phages have been implicated in easing disease burden, particularly in the endemic region surrounding the Bay of Bengal[12,13]. We recently described the isolation of the ICP1 (for the International Centre for Diarrhoeal Disease Research, Bangladesh cholera phage 1)-related, *V. cholerae* O1-specific virulent myoviruses that are omnipresent among cholera patient rice-water stool samples collected at the ICDDR,B from 2001 to 2011 (ref. 14 and present study). *V. cholerae* readily evolves resistance to ICP1 predation through mutations in O1 antigen biosynthetic genes outside the human host; however, this mutational escape comes at a cost as virulence necessitates maintenance of the O1 antigen[15]. This dynamic between predation by ICP1 and virulence of *V. cholerae* O1, specifically in the context of human infection, provides a unique opportunity for discovery of novel bacterial immunity and phage immune evasion strategies. One bacterial defensive strategy against phages is the CRISPR/Cas system. CRISPR loci consist of an array of short direct repeats separated by highly variable spacer sequences of precise length corresponding to segments of previously captured foreign DNA (protospacers)[4,7,9]. CRISPR loci are found in ~40% and ~90% of sequenced bacterial and archaeal genomes, respectively[8,16]. The CRISPR array is transcribed and the transcript cleaved into small CRISPR RNAs (crRNAs) that, in conjunction with the Cas proteins, execute an efficient process of immunity in which foreign nucleic acids are recognized by hybridization to crRNAs and cleaved[4,7,8].

We isolated eleven ICP1-related phages from stools of cholera patients at the ICDDR,B (ref. 14 and present study), five of which encode a CRISPR/Cas system located between open reading frames (ORFs) 87 and 88 of the ancestral ICP1 genome[14]. The GC content of this CRISPR/Cas system is the same (~37%) as the rest of the ICP1 genome. The ICP1 CRISPR/Cas system consists of two CRISPR loci (designated CR1 and CR2) and six *cas* genes (Fig. 1a) whose organization and protein products are most homologous to Cas proteins of the type 1-F (*Yersinia pestis*) subtype system[17] (Supplementary Table 1). *V. cholerae* is divided into two biotypes, classical and El Tor, the former of which is associated with earlier pandemics and has since been replaced by the El Tor biotype[18]. The classical strain, *V. cholerae* O395, has a CRISPR/Cas system belonging to the type I-E (*Escherichia coli*) subtype[17], and to date there has not been any description of El Tor strains possessing a CRISPR/Cas system. Thus, the origin of the CRISPR/Cas system in ICP1 phage is unknown. Protospacer-adjacent motifs (PAMs) are type-specific, short conserved sequence motifs in the immediate vicinity of protospacers that are required for acquisition and targeting[7,9,11,19]. In contrast to the GG PAM reported for the type I-F CRISPR/Cas systems in bacteria[19], the protospacers targeted by the ICP1 CRISPR array have a GA PAM (Supplementary Fig. 1).

The majority of spacers in the ICP1 CRISPR show 100% identity to sequences within an 18-kilobase (kb) island found in a subset of *V. cholerae* strains that include the classical strain O395 isolated in India in 1964, El Tor strain MJ-1236 isolated in Bangladesh in 1994, and several El Tor strains collected at the ICDDR,B between 2001 and 2011 (Supplementary Table 2). The 18-kb island resembles the phage-inducible chromosomal islands (PICIs) of Gram-positive bacteria, including the prototype *Staphylococcus aureus* pathogenicity islands (SaPIs)[20,21]. SaPIs are induced to excise, circularize and replicate

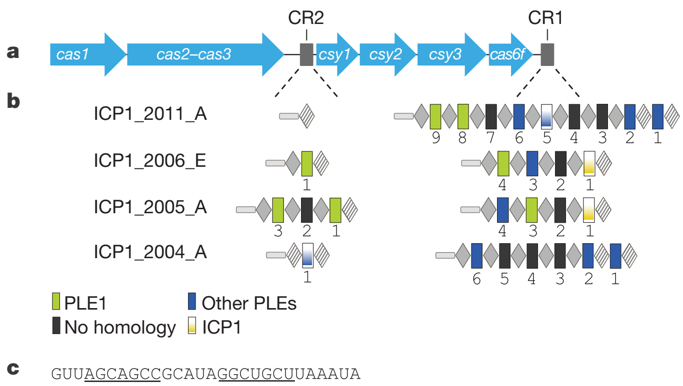


Figure 1 | Genomic organization of the ICP1 CRISPR/Cas system. **a**, The ICP1 phage CRISPR/Cas system consists of six *cas* genes and two CRISPR loci (CR1 and CR2). **b**, For each CRISPR locus, the repeat (28 bp) and spacer (32 bp) content is detailed as grey diamonds and coloured rectangles, respectively. Repeats (28 bp) that match the repeat consensus are shown in grey diamonds, and degenerate repeats are indicated in hatched grey diamonds. An AT-rich leader sequence precedes each CRISPR locus (grey rectangle). Spacers are coloured according to the percentage identity (solid represent 100% identity, gradient represents 81–97% identity). A fifth ICP1-related phage (ICP1_2003_A) has a genetically identical CRISPR/Cas system to ICP1_2004_A, and has been omitted for simplicity. **c**, The RNA sequence of the CR1 and CR2 consensus repeat with the partially palindromic sequence forming the predicted stem in the crRNA underlined.

[1]Howard Hughes Medical Institute and Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, Massachusetts 02111, USA. [2]Division of Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [3]Harvard Medical School, Boston, Massachusetts 02114, USA.

following infection by certain phages. They use varied mechanisms to interfere with the phage reproduction cycle to enable their own promiscuous spread[21], and this can protect the surrounding bacterial population from further phage predation. The organization of the *V. cholerae* 18-kb island targeted by the ICP1 CRISPR/Cas system is similar in length, base composition and organization to that observed in the SaPIs subset of PICIs, with an integrase homologue at one end and a GC content lower than that of the host species (37% compared to 47.5%). We therefore refer to the 18-kb element as the *V. cholerae* PICI-like element (PLE) (Fig. 2).

To address the functional relevance of the ICP1 CRISPR/Cas system, we focused on the interaction between the paired ICP1_2011_A phage and the *V. cholerae* O1 El Tor strain (harbouring PLE1) that were isolated from the same stool sample (for simplicity hereafter referred to as ICP1 and *V. cholerae* PLE$^+$). ICP1 has two CRISPR spacers (8 and 9) (Fig. 1b) that have 100% identity to sequences within the *V. cholerae* PLE (Fig. 2 and Supplementary Table 2). Using the standard soft agar overlay method, we found that ICP1 can plaque efficiently on *V. cholerae* PLE$^+$ (Fig. 3b). We used northern blot analysis to confirm that ICP1 crRNAs are transcribed and processed during *V. cholerae* infection (Supplementary Fig. 2). To test whether targeting of the PLE by the ICP1 CRISPR/Cas system affects phage fitness, we eliminated spacer 8 and 9 targeting. Spacer 8 targeting was disrupted by introducing silent mutations into its target within the PLE, generating *V. cholerae* PLE(8*) (Fig. 3a). We then infected this strain with a spontaneous ICP1 spacer 9 deletion mutant, referred to as ICP1(ΔS9). ICP1(ΔS9) was blocked for plaque formation on *V. cholerae* PLE(8*); however, it maintained wild-type plaquing efficiency on *V. cholerae* PLE$^+$ (Fig. 3b). Importantly, *V. cholerae* PLE(8*) is sensitive to plaque formation by ICP1 (Fig. 3b), which still harbours one spacer (S9) targeting the PLE. These results demonstrate that ICP1 CRISPR/Cas must target the PLE for destruction in order to effectively infect and form plaques, and that a single spacer that targets the PLE is sufficient to facilitate successful phage replication. A mutant in which PLE ORFs 7–20 were deleted was susceptible to infection by ICP1(ΔS9) with wild-type plaquing efficiency (Supplementary Fig. 3). This demonstrates that an intact PLE is required to inhibit ICP1 in the absence of CRISPR targeting. These results, in conjunction with the observation that PLE1 circularizes following ICP1 infection (Supplementary Fig. 4), further support our designation of the 18-kb island as a PICI-like element.

It has been well documented in the type I-E (*E. coli*) system that CRISPR interference requires an intact PAM and a fully complementary seed region (a non-contiguous 7 base pair (bp) sequence immediately adjacent to the PAM)[22]. To address the sequence requirements of the ICP1 CRISPR/Cas system we constructed a series of point mutations in the spacer 8 target in *V. cholerae* PLE that span the PAM, seed region and remainder of the target sequence, and determined their effect on immunity. In accordance with previous results, we found that single mutations within the PAM or the first four positions in the seed region immediately adjacent to the PAM abolish ICP1 CRISPR/Cas immunity (Supplementary Fig. 5). Interestingly, mutations of increasing distance from the PAM showed a concordant decreasing effect on immunity. Up to five mismatches outside of the seed region of the target are known to be tolerated in the type I-E system[22], and similarly we found that three and five mutations outside of the seed region were tolerated; however, eight mutations were not (Supplementary Fig. 5).

a

```
V. cholerae PLE(WT)  AAT TTA AAT AGG GAA GAT AAG CAA AGG GTT GAC
                      N   L   N   R   E   D   K   Q   R   V   D
V. cholerae PLE(8*)  AAC TTG AAC AGA GAG GAC AAA CAG AGA GTC GAT
```

b

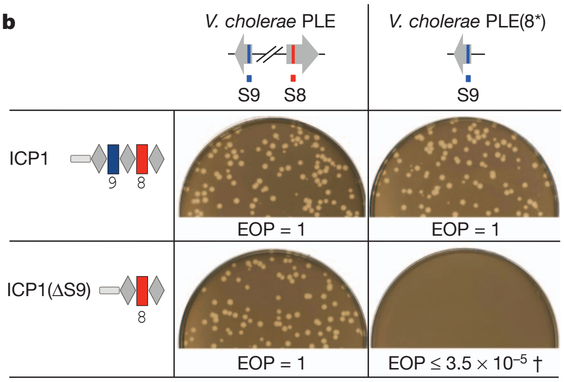


Figure 3 | Sequence-based targeting by the ICP1 CRISPR/Cas system is essential for lytic growth on *V. cholerae* PLE$^+$. **a**, Disruption of the *V. cholerae* PLE target protospacer generating *V. cholerae* PLE(8*). The 32 bp protospacer sequence is shaded in grey. **b**, The sensitivity of each strain (top row) to ICP1 or ICP1(ΔS9) (left column) is shown. Identity between the spacer and targeted protospacer is indicated by the red and blue rectangles. The efficiency of plaquing (EOP, which is the plaque count on the mutant host strain divided by that on the wild-type host strain) is indicated. A dagger indicates that the EOP is 10^{-5} or 10^{-8} depending on the presence of PLE in the host strain used for propagation as discussed in the text.

In experiments where the ICP1 CRISPR/Cas system could not target the *V. cholerae* PLE and therefore plaque formation was greatly reduced, we observed phage escape mutants at frequencies that were dependent on the host strain on which the phage had been previously propagated. When ICP1(ΔS9) was grown on a PLE$^+$ host before plaquing on *V. cholerae* PLE(8*), the efficiency of plaquing (EOP, which is the plaque count on the mutant host strain divided by that on the wild-type host strain) was 3.5×10^{-5}. The CRISPR loci from ten independent ICP1(ΔS9) escape mutants were sequenced, and in all cases, a new spacer was present at the leader end of the CRISPR CR1 array. Furthermore, the new spacers had 100% identity to sequences within the PLE (Fig. 2), and all newly integrated spacers target the PLE with the conserved GA dinucleotide PAM sequence (Supplementary Fig. 1b). The experimentally acquired spacers target both the coding and noncoding strands (Supplementary Table 3), although most (nine out of ten) target the coding strand. The pre-existing spacer (S8) (although mismatched in these experiments) also targets the coding strand; these data are in support of recent evidence that the DNA strand from which new protospacers are incorporated is heavily biased towards the existing protospacer orientation[23,24]. In contrast to when phage were propagated on a PLE$^+$ host before plaquing on *V. cholerae* PLE(8*), phage escape mutants were detected at a much lower frequency (EOP $= 1.1 \times 10^{-8}$) when ICP1(ΔS9) was grown on a *V. cholerae* PLE$^-$ host. This shows that new spacers targeting the PLE are incorporated into the CRISPR array during ICP1(ΔS9) infection of the PLE$^+$ host (the immunization process), and that an immune host possessing an untargeted PLE can subsequently be used to select for new ICP1 CRISPR acquisition events that confer targeting and thus restore phage replication. These results demonstrate that the ICP1

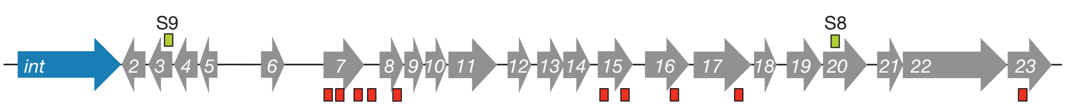


Figure 2 | Genomic organization of PLE1, a representative *V. cholerae* PLE targeted by the CRISPR/Cas system of ICP1-related phages. The integrase (*int*) is in blue, genes encoding hypothetical proteins (with numerical ORF designations) are grey. The locations of protospacers incorporated into the CRISPR locus as spacers 8 and 9 (S8 and S9 of ICP1_2011_A) are indicated in green above the map. The locations of experimentally acquired protospacers are shown below the map in red.

CRISPR/Cas system is fully functional as an adaptive immune evasion system that benefits the phage.

ICP1 has evolved to effectively target the *V. cholerae* PLE with an adaptive immune evasion system that has never before been shown to function in bacterial viruses. During ICP1 infection of *V. cholerae* PLE^{+}, PLE circularizes (Supplementary Fig. 4) and inhibits ICP1 through an unknown mechanism. To replicate successfully, ICP1 uses the CRISPR/Cas system to target the PLE for destruction. Because host cell death and DNA damage is inherent to lytic phage infection, CRISPR-mediated DNA cleavage of the PLE does not affect ICP1 infection. Sequencing data has been used to identify putative CRISPR arrays within a *Clostridium difficile* prophage[25], and more recently in metagenomic data sets of free viruses[26,27]. However, there is currently no evidence for expression or function of these putative arrays. We show that the ICP1-encoded CRISPR/Cas system actively and autonomously functions to inhibit host immunity and thereby permit lytic infection. This finding, in conjunction with the previous observations regarding the presence of CRISPR loci in other phages[25–27], suggests that the use of the so-called bacterial adaptive immune system by these bacterial predators may be an underappreciated immune evasion strategy in the unfolding phage versus host co-evolutionary arms race.

METHODS

Phages (ICP1_2011_A and ICP1_2006_E) and *V. cholerae* were isolated from cholera rice-water stool samples and propagated as described[14,15]. Genomic libraries were generated for phage and host strains as described[28] and sequenced using an Illumina HiSeq2000. A *V. cholerae* O1 El Tor isolate collected at the ICDDR,B in 2006, which was sequenced in this study and found to not harbour a PLE, was used as the PLE^{-} host for propagation experiments. We used the CRISPRFinder program[16] to identify CRISPR loci. WebLogo[29] was used to generate sequence logos for identification of the PAM. Point mutations were constructed using splicing by overlap extension (SOE) PCR and introduced using pCVD442-lac as previously described[15]. The PLE1 deletion construct (missing 8.6 kb including ORFs 7–20) was constructed using SOE PCR and introduced by natural transformation with subsequent deletion of the antibiotic-resistance marker using the FLP recombinase method as described[30]. ICP1(ΔS9) was identified by screening for alterations in the CRISPR array by PCR following growth on *V. cholerae* PLE^{+}. RNA was purified using the mirVana kit (Ambion) at the indicated times and run on 12% polyacrylamide urea gels. Northern blots were pre-hybridized in Ultrahyb-oligo (Ambion) and hybridization was carried out at 37 °C overnight using 32-nucleotide 5′ end-labelled DNA probes (generated with [γ-32P]ATP and T4 polynucleotide kinase) complementary to spacers 8 and 6.

Received 9 November 2012; accepted 23 January 2013.

1. Brüssow, H. & Hendrix, R. W. Phage genomics: small is beautiful. *Cell* **108,** 13–16 (2002).
2. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature Rev. Microbiol.* **8,** 317–327 (2010).
3. Stern, A. & Sorek, R. The phage-host arms race: shaping the evolution of microbes. *Bioessays* **33,** 43–51 (2011).
4. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327,** 167–170 (2010).
5. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).
6. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322,** 1843–1845 (2008).
7. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* **45,** 273–297 (2011).
8. van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M. & Brouns, S. J. J. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* **34,** 401–407 (2009).
9. Karginov, F. V. & Hannon, G. J. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell* **37,** 7–19 (2010).
10. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).
11. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Rev. Genet.* **11,** 181–190 (2010).
12. Faruque, S. M. *et al.* Self-limiting nature of seasonal cholera epidemics: role of host-mediated amplification of phage. *Proc. Natl Acad. Sci. USA* **102,** 6119–6124 (2005).
13. Nelson, E. J., Harris, J. B., Morris, J. G., Calderwood, S. B. & Camilli, A. Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nature Rev. Microbiol.* **7,** 693–702 (2009).
14. Seed, K. D. *et al.* Evidence of a dominant lineage of *Vibrio cholerae*-specific lytic bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh. *MBio* **2,** e00334–10 (2011).
15. Seed, K. D. *et al.* Phase variable O antigen biosynthetic genes control expression of the major protective antigen and bacteriophage receptor in *Vibrio cholerae* O1. *PLoS Pathog.* **8,** e1002917 (2012).
16. Grissa, I., Vergnaud, G. & Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8,** 172 (2007).
17. Makarova, K. S. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nature Rev. Microbiol.* **9,** 467–477 (2011).
18. Longini, I. M. *et al.* Epidemic and endemic cholera trends over a 33-year period in Bangladesh. *J. Infect. Dis.* **186,** 246–251 (2002).
19. Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155,** 733–740 (2009).
20. Novick, R. P., Christie, G. E. & Penadés, J. R. The phage-related chromosomal islands of Gram-positive bacteria. *Nature Rev. Microbiol.* **8,** 541–551 (2010).
21. Ram, G. *et al.* Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc. Natl Acad. Sci. USA* **109,** 16300–16305 (2012).
22. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl Acad. Sci. USA* **108,** 10098–10103 (2011).
23. Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Commun.* **3,** 945–947 (2012).
24. Swarts, D. C., Mosterd, C., van Passel, M. W. J. & Brouns, S. J. J. CRISPR interference directs strand specific spacer acquisition. *PLoS ONE* **7,** e35888 (2012).
25. Sebaihia, M. *et al.* The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nature Genet.* **38,** 779–786 (2006).
26. Minot, S. *et al.* The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* **21,** 1616–1625 (2011).
27. Garcia-Heredia, I. *et al.* Reconstructing viral genomes from the environment using fosmid clones: the case of haloviruses. *PLoS ONE* **7,** e33802 (2012).
28. Lazinski, D. W. L. & Camilli, A. Homopolymer tail-mediated ligation PCR: a streamlined and highly efficient method for DNA cloning and library construction. *Biotechniques* **54,** 25–34 (2013).
29. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).
30. De Souza Silva, O. & Blokesch, M. Genetic manipulation of *Vibrio cholerae* by combining natural transformation with FLP recombination. *Plasmid* **64,** 186–195 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank the Tufts University Core Facility for sequencing and computational support. This work was supported by US National Institutes of Health grants AI055058 (A.C.), AI045746 (A.C.) and AI058935 (S.B.C.). A.C. is a Howard Hughes Medical Institute Investigator.

Author Contributions K.D.S. and D.W.L. performed experiments. K.D.S., D.W.L. and A.C. designed experiments. K.D.S. and A.C. wrote the manuscript. S.B.C. provided materials. All authors discussed the results and commented on the manuscript.

Author Information The sequences for the CRISPR/Cas system in ICP1_2011_A and ICP1_2006_E have been deposited at GenBank (accession numbers KC152959 and KC152958, respectively). The sequences for the *V. cholerae* PLEs identified in clinical isolates from the ICDDR,B have been deposited at GenBank/EMBL/DDBJ under the accession numbers KC152960 (PLE1) and KC152961 (PLE2). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.C. (andrew.camilli@tufts.edu).

nature biotechnology

Photoactivatable CRISPR-Cas9 for optogenetic genome editing

Yuta Nihongaki, Fuun Kawano, Takahiro Nakajima & Moritoshi Sato

We describe an engineered photoactivatable Cas9 (paCas9) that enables optogenetic control of CRISPR-Cas9 genome editing in human cells. paCas9 consists of split Cas9 fragments and photoinducible dimerization domains named Magnets. In response to blue light irradiation, paCas9 expressed in human embryonic kidney 293T cells induces targeted genome sequence modifications through both nonhomologous end joining and homology-directed repair pathways. Genome editing activity can be switched off simply by extinguishing the light. We also demonstrate activation of paCas9 in spatial patterns determined by the sites of irradiation. Optogenetic control of targeted genome editing should facilitate improved understanding of complex gene networks and could prove useful in biomedical applications.

The type II bacterial clustered, regularly interspaced, short palindromic repeats (CRISPR) and the CRISPR-associated protein 9 (Cas9), known as CRISPR-Cas9, mediates targeted genome modifications that enable dissection of gene and regulatory functions[1–3]. The *Streptococcus pyogenes* Cas9 nuclease (hereafter referred to as Cas9) can bind to and cleave a target DNA sequence that is complementary to the first 20 nucleotides of a single-guide RNA (sgRNA) and is adjacent to a protospacer-adjacent motif (PAM) of the form NGG. Cas9-induced DNA double-strand breaks are repaired by nonhomologous end joining (NHEJ) or homology-directed repair (HDR) in mammalian cells, thereby enabling targeted genome editing.

Methods have been developed to chemically control the nuclease activity of Cas9, such as doxycycline-regulated Cas9 expression[4,5], rapamycin-inducible split-Cas9 (ref. 6) and transient delivery of purified Cas9:sgRNA complex[7–9]. These chemical methods have been used for generating conditional gene knockouts and reducing levels of off-target genome modification. However, some chemicals have adverse effects. For example, rapamycin can induce undesirable biological effects by perturbing the endogenous mammalian target of rapamycin (mTOR) pathway[10]. Also, because chemicals diffuse freely and are difficult to rapidly remove, such methods cannot be applied to achieve spatiotemporal genome editing.

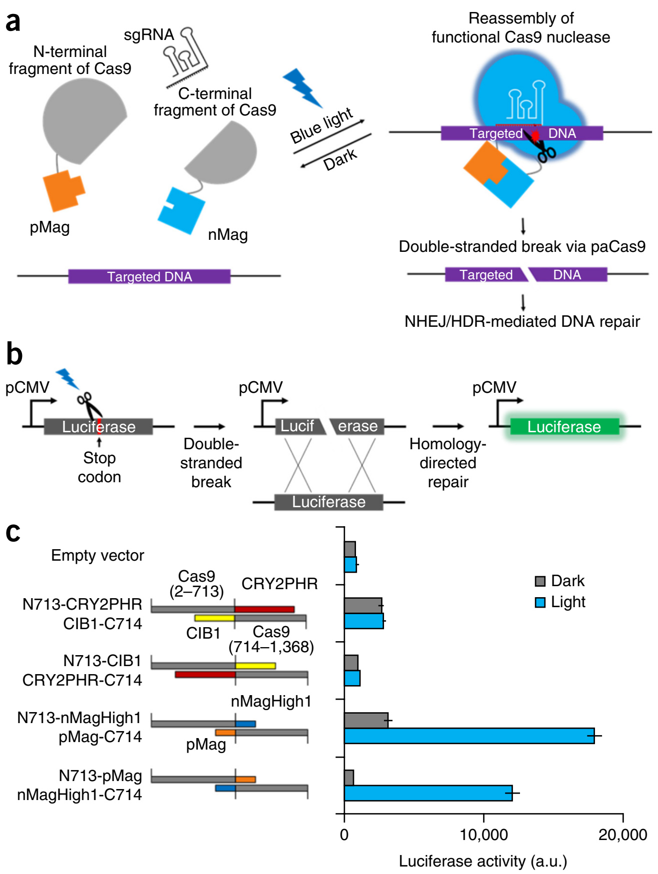
We set out to design a method of controlling Cas9 nuclease activity that is noninvasive and incorporates spatial, temporal and reversible control. Light has high spatiotemporal resolution and is noninvasive[11,12] but methods to optically control Cas9 nuclease activity have so far been elusive.

To produce an optically controlled Cas9 we fused two split Cas9 fragments with photoinducible dimerization domains to generate paCas9 (**Fig. 1a**). In initial experiments to determine the best suited split site of Cas9 for high efficiency inducible control, we generated various Cas9 fragments fused with the rapamycin-inducible dimerization system, FKBP-FRB[13] (**Supplementary Fig. 1**). We selected 18 candidate split sites based on an analysis of the crystal structure of Cas9 in complex with sgRNA[14,15]. All candidate split site positions were loop regions exposed to solvent. We assessed the rapamycin-induced nuclease activity of each split-Cas9 pair using a luciferase-reporter plasmid HDR assay (**Fig. 1b**). In this assay, the cytomegalovirus (CMV) promoter-driven luciferase reporter with an in-frame stop codon (StopFluc-1) is cleaved by split-Cas9, and then recovers full-length luciferase expression through homologous recombination with promoter-less luciferase donor vector. Eight combinations of N- and C-terminal Cas9 fragments showed significant rapamycin-induced reporter upregulation in HEK293T cells (**Supplementary Fig. 2**). In subsequent experiments, we used the N-terminal fragment of Cas9 (residues 2–713, named N713) and C-terminal fragment of Cas9 (residues 714–1,368, named C714), which was one of the most effective rapamycin-inducible split-Cas9 pairs.

Next, we fused photoinducible dimerization domains with N713 and C714 (**Fig. 1c**). First, we tested the CRY2-CIB1 photoinducible dimerization system, which is based on blue light–dependent protein interactions between *Arabidopsis thaliana* cryptochrome 2 (CRY2) and its binding partner CIB1 (ref. 16). This system has been widely used for optogenetic control of protein-protein interactions in mammalian cells. We generated N713 and C714 fused to the photolyase homology region of CRY2 (CRY2PHR) and CIB1 and tested induction potency using a luciferase plasmid HDR assay. However, N713 and C714 fused to the CRY2PHR and CIB1 system did not show light-induced Cas9 activity. There are several possible explanations for this failure. First, steric hindrance could be caused by CRY2PHR (498 amino acids) and CIB1 (335 amino acids), impeding reassembly of split-Cas9. Second, oligomerization of CRY2PHR might reduce or preclude interactions between the N-terminal and C-terminal fragments of Cas9 (ref. 17). Keeping these potential problems in mind we next focused on a recently developed photoinducible dimerization system named Magnets[18]. The Magnet system consists of paired photoswitchable proteins, named positive Magnet (pMag) and negative Magnet (nMag). Upon blue light irradiation, pMag and nMag heterodimerize. Unlike CRY2-CIB1, pMag and nMag (150 amino

Graduate School of Arts and Sciences, The University of Tokyo, Komaba, Meguro-ku, Tokyo, Japan. Correspondence should be addressed to M.S. (cmsato@mail.ecc.u-tokyo.ac.jp).

Received 21 February; accepted 5 May; published online 15 June 2015; doi:10.1038/nbt.3245



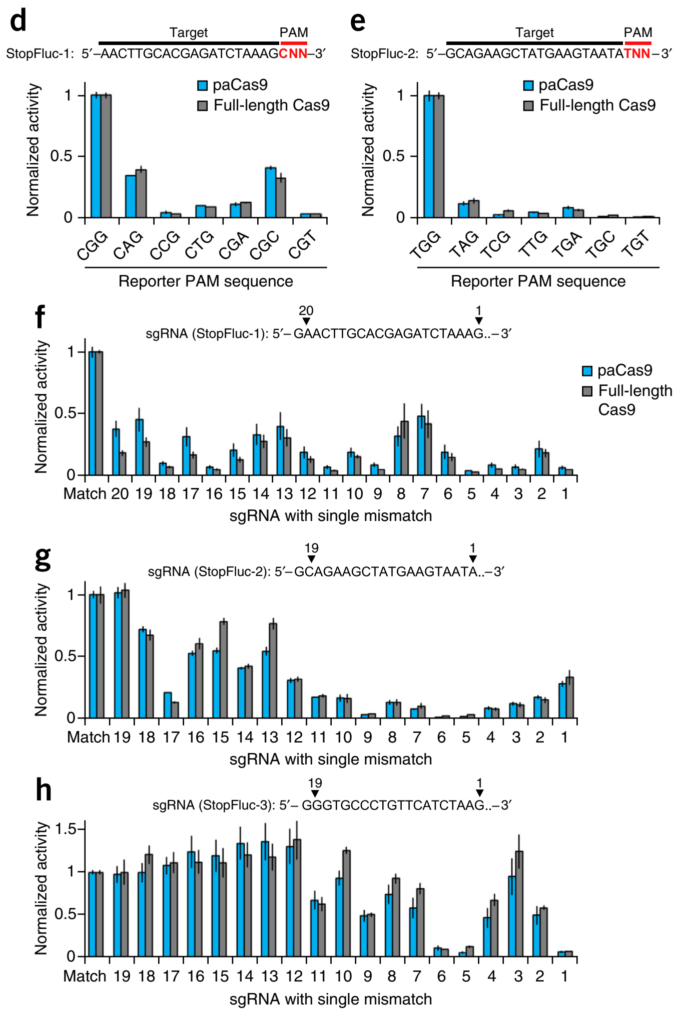


Figure 1 Design and characterization of photoactivatable Cas9. (**a**) Schematic of the photoactivatable Cas9 (paCas9). Cas9 is split into two fragments without nuclease activity, and the Cas9 fragments are fused with photoinducible dimerization domains (pMag and nMag). Blue light irradiation induces heterodimerization between pMag and nMag, which enables split Cas9 fragments to reassociate, thereby reconstituting RNA-guided nuclease activity. Gray, inactive; blue, active. (**b**) Luciferase reporter plasmid HDR assay. When Cas9 cleaves the CMV-driven luciferase reporter with an in-frame stop codon (StopFluc-1), the luciferase reporter is repaired by homologous recombination with promoter-less luciferase donor vector and recovers bioluminescence activity. (**c**) Light-induced reporter activity in HEK293T cells using N713 and C714 fragments of Cas9 fused with photoinducible dimerization domains. (**d**,**e**) Activities of paCas9-1 and full-length Cas9 targeting StopFluc-1 (**d**) and StopFluc-2 (**e**), harboring indicated mutations in the PAM. Values are normalized to a positive control, which is a luciferase reporter with canonical PAM (NGG). (**f**–**h**) Activities of Cas9 and paCas9 targeting StopFluc-1 (**f**), StopFluc-2 (**g**) and StopFluc-3 (**h**) with a set of sgRNAs harboring single-nucleotide Watson-Crick transversion mutations. sgRNAs mutated in 5′-end G to C were not tested because 5′-end G is necessary for efficient expression from the U6 promoter. The positions of point mutations in each sgRNA are indicated at the top of each panel. Values are normalized to positive control with perfectly matched sgRNAs. In (**c**–**h**), data are shown as the mean ± s.e.m. (n = 6 from two individual experiments with biological triplicates that is, separately seeded cells).

acids each) are of a similar size to FKBP (107 amino acids) and FRB (93 amino acids). The dynamic range and dissociation kinetics of the Magnet system can be tuned by mutating pMag and/or nMag. We used pMag and nMagHigh1 (nMag with M135I and M165I mutations)[18]. To test whether Magnets could provide effective light-triggered reassembly of a split-Cas9, we tested two pairs of fusion proteins; N713-nMagHigh1 and pMag-C714 and N713-pMag and nMagHigh1-C714. We found that both paired fusion proteins showed substantial light-induced Cas9 activity, and N713-pMag and nMagHigh1-C714 yielded the highest fold-induction (16.4-fold) and lowest background activity (**Fig. 1c**). We used N713-pMag and nMagHigh1-C714 and named this construct paCas9-1 (**Supplementary Fig. 3**).

To investigate whether paCas9-1 recognizes PAM in the same way as full-length Cas9, we generated a set of luciferase reporters, into which we inserted stop codons, and which harbored a point mutation in NGG PAM (**Fig. 1d**,**e**). We tested two different luciferase reporters containing an internal stop codon in different sites (StopFluc-1 and StopFluc-2) and confirmed that the Cas9-induced activities of the luciferase reporters that had noncanonical PAM were lower than that of luciferase reporters having the canonical PAM of the form NGG. We found no significant difference in the normalized luciferase activities of paCas9-1 and full-length Cas9. We also evaluated the DNA targeting specificity of paCas9-1 using a luciferase plasmid HDR assay (**Fig. 1f**). We produced a series of sgRNAs for StopFluc-1 harboring single-nucleotide Watson-Crick transversion mutations (**Supplementary Table 1**). We found no significant difference in the DNA targeting specificities of paCas9-1 and full-length Cas9. To further investigate DNA targeting specificity of paCas9-1, we performed the specificity assay using two different reporters containing an internal stop codon in different sites (StopFluc-2 and StopFluc-3) and confirmed that the DNA targeting specificity of paCas9-1 is comparable with that of full-length Cas9 (**Fig. 1g**,**h**). Consistent with previous

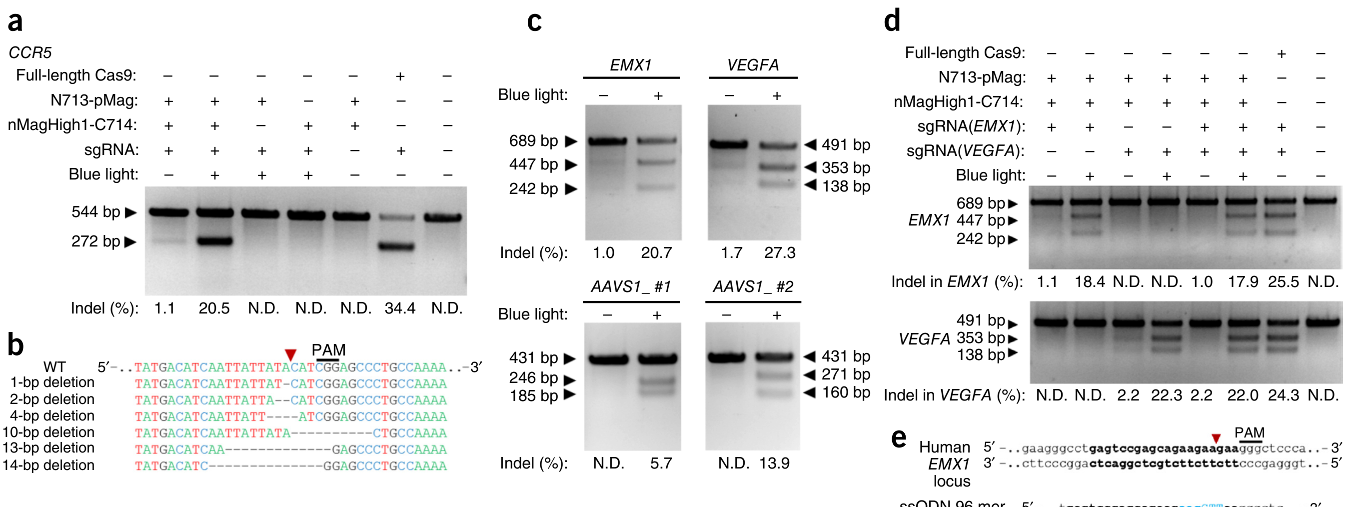


Figure 2 Optogenetic genome editing of mammalian endogenous genes by the photoactivatable Cas9. (**a**) Light-mediated indel mutation of human *CCR5* locus by paCas9-1. Indel mutation frequencies were evaluated by mismatch-sensitive T7E1 assay. (**b**) Example sequences of the human *CCR5* locus targeted by paCas9. Six clones harbored deletion mutations (assessed by Sanger sequencing of 27 individual cDNA clones). (**c**) paCas9 was used to target a variety of endogenous genes. Cells were transfected with paCas9-1 and sgRNA targeting the *EMX1* site, the *VEGFA* site and the two sites in *AAVS1*. (**d**) Light-induced multiplex genome editing by paCas9. HEK293T cells were transfected with paCas9-1 and the indicated sgRNAs. (**e**) Schematic of paCas9-mediated precise genome editing experiments. The red arrowhead indicates the putative paCas9 cleavage site. A 96-mer, single-stranded oligodeoxynucleotide (ssODN) donor template is designed to insert HindIII site (blue) into *EMX1* locus. Target sequence is boldface lowercase. 3-base-pair insertion sequence is uppercase. (**f**) paCas9-mediated precise genome editing by HDR by means of a single-stranded oligodeoxynucleotide template. Successful HDR frequency was determined by restriction-fragment length polymorphism (RFLP) analysis, and calculated as the ratio of HindIII digested product to substrate. NHEJ and HDR rates presented are means N.D., not determined. (In **a**, $n = 3$ from independent experiments. In **c**,**d** and **f**, $n = 2$ from two independent experiments.) In **a**,**c**,**d**,**f**, 20 h after transfection, samples were illuminated by 1.2 W/m^2 blue light or kept in the dark for 24 h (**a**,**c**,**d**) or 48 h (**f**) before genome extraction. Full-length gels are presented in **Supplementary Figure 8**.

studies, sensitivity patterns to single sgRNA-DNA mismatch differed among target sequences[19–21]. From these experiments, we conclude that the PAM requirement and target specificities of paCas9-1 are indistinguishable from those of full-length Cas9.

To show that paCas9-1 could cleave a targeted endogenous genomic locus in mammalian cells and induce indel mutation by nonhomologous end joining (NHEJ) in a light-dependent fashion, we transfected HEK293T cells with paCas9-1 and an sgRNA targeting the human *CCR5* locus (**Fig. 2a**). We quantitatively evaluated the ability of paCas9-1 to induce indel mutations in response to light using the mismatch-sensitive T7 endonuclease I (T7E1), which cleaves heteroduplexed DNA strands formed by hybridization between mutant and wild-type DNA. In the dark, cells transfected with paCas9-1 targeting *CCR5* showed only 1.1% indel rates. However, upon blue light irradiation, cells transfected with paCas9-1 targeting *CCR5* exhibited significantly ($P = 0.012$) higher indel rates (20.5%) in the human *CCR5* locus. The frequency of indel mutations induced by paCas9-1 is ~60% of that achievable with full-length Cas9 (paCas9-1: 20.5%, full-length Cas9: 34.4%) (**Fig. 2a**). Using Sanger sequencing, we verified that paCas9-1–mediated indel mutations occurred in the targeted region of the human *CCR5* locus (**Fig. 2b**). To explore the generalizability of optogenetic RNA-guided genome editing with paCas9-1, we used sgRNAs for four additional sites in three human genes (*EMX1*, *VEGFA* and *AAVS1*). With each of these sgRNAs, we observed light-induced indel mutations were observed (**Fig. 2c**). We also carried out a time-course analysis of indel mutations in the *EMX1* locus induced by paCas9-1 (**Supplementary Fig. 4**). We observed that the frequency of indel mutations increased as the blue light irradiation time increased. To test whether paCas9-1 could induce indel mutations in different cell lines, we transfected HeLa cells with paCas9-1 and an sgRNA targeting human *EMX1* (**Supplementary Fig. 5**). Light-induced indel mutations in the *EMX1* locus were also observed in HeLa cells. We also tested whether paCas9-1 could induce indel mutations at multiple target sites (**Fig. 2d**). Using two sgRNAs targeting *EMX1* and *VEGFA* simultaneously, paCas9-1 induced indel mutations in both human *EMX1* and *VEGFA* loci in response to light. These results show that paCas9-1 can be deployed for optogenetic multiplexed control of NHEJ-mediated indel mutations in mammalian cells.

Next, we investigated whether paCas9-1 could be used for genome editing by means of HDR (**Fig. 2e**,**f**). We used single-stranded oligodeoxynucleotides (ssODN) as a donor template. We transfected HEK293T cells with paCas9-1 targeting *EMX1* and ssODN containing a HindIII site and analyzed the frequency of HDR in the EMX1 locus using a restriction fragment length polymorphism (RFLP) assay. We found that paCas9-1 induces HindIII site integration in the human *EMX1* locus at a frequency of 7.2%. This shows that paCas9-1 can induce both random indel mutations and designed genome sequence modification through HDR in response to light.

The off-target activity of Cas9 has been reduced using a paired nicking strategy based on the Cas9 D10A variant, which nicks targeted DNA instead of cutting double strands[22]. To explore whether paCas9-1 could be converted into a photoactivatable nickase, we produced paCas9-1 containing the D10A mutation (paCas9 nickase) (**Supplementary Fig. 6**). Using the paCas9 nickase with an sgRNA targeting *EMX1* did not induce indel mutations in the human *EMX1* locus. However, using a pair of sgRNAs targeting opposite strands of the *EMX1* site, our paCas9 nickase induced indel mutations in

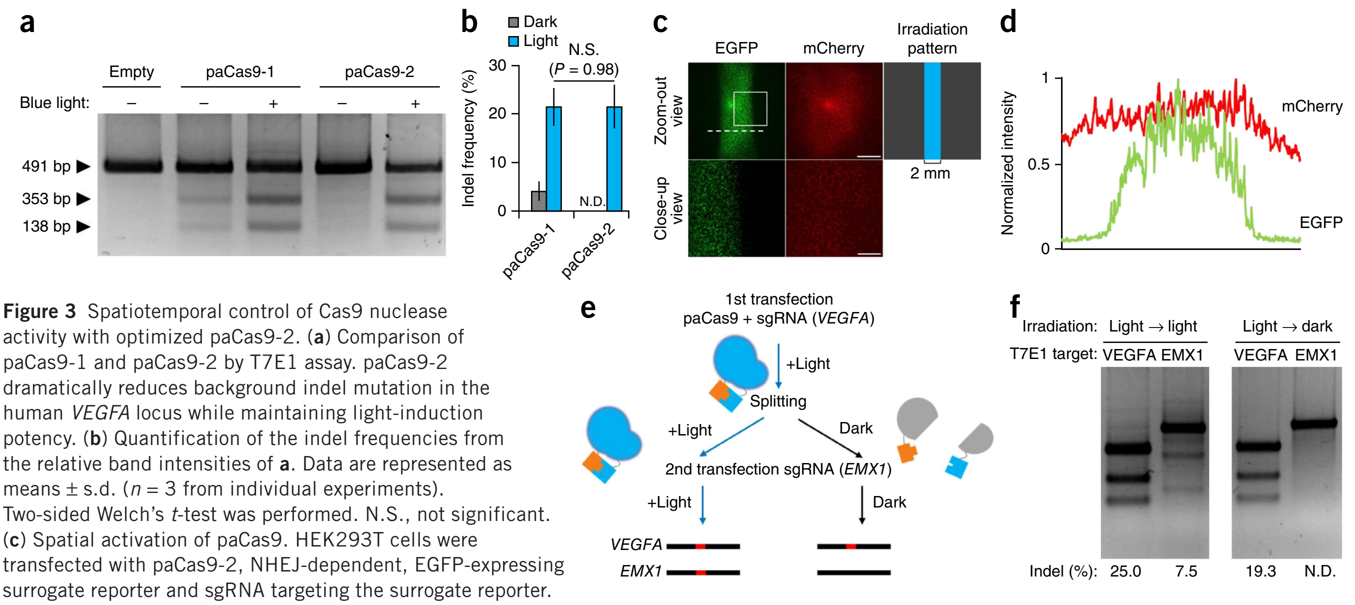


Figure 3 Spatiotemporal control of Cas9 nuclease activity with optimized paCas9-2. (**a**) Comparison of paCas9-1 and paCas9-2 by T7E1 assay. paCas9-2 dramatically reduces background indel mutation in the human *VEGFA* locus while maintaining light-induction potency. (**b**) Quantification of the indel frequencies from the relative band intensities of **a**. Data are represented as means ± s.d. ($n = 3$ from individual experiments). Two-sided Welch's *t*-test was performed. N.S., not significant. (**c**) Spatial activation of paCas9. HEK293T cells were transfected with paCas9-2, NHEJ-dependent, EGFP-expressing surrogate reporter and sgRNA targeting the surrogate reporter. 20 h after transfection, samples were irradiated by slit-patterned blue light using a photomask for 24 h. The width of the slit is 2 mm. Close-up view within white square region is also shown. Scale bar, 3 mm (zoom-out view) and 1 mm (close-up view), respectively. (**d**) Line scan intensity profile of EGFP (green) and mCherry (red) in **c**. (**e**) Experimental scheme to test whether paCas9 activation is reversible. First, HEK293T cells were transfected with paCas9-2 and sgRNA targeting *VEGFA*. After 20 h, cells were illuminated with blue light for 6 h, then split and incubated in the light or the dark. After 6 h incubation, the cells were transfected with an sgRNA targeting *EMX1*, and incubated again in the light or the dark. 30 h later, genomic DNA was extracted. If paCas9 activation is reversible, the cells shifted to the dark state before the second transfection of sgRNA targeting *EMX1* should show the indel mutation in only the *VEGFA* locus. (**f**) A representative gel of the T7E1 assay in **e**. The indel frequencies of the *EMX1* and *VEGFA* locus are shown at the bottom of the gel. ($n = 4$ from two independent experiments with biological duplicates). Full-length gels are presented in **Supplementary Figure 8**.

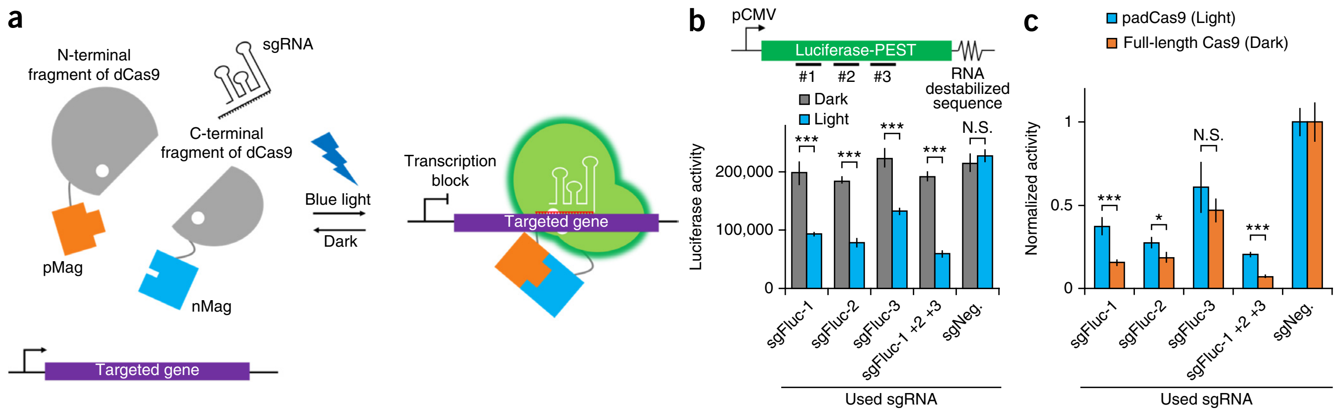
response to light. This indicates that the double-nicking strategy can be applied to paCas9 to reduce off-target genome modifications.

Although paCas9-1 can induce efficient NHEJ-mediated indel mutation, paCas9-1 had background activity (about 1–3%) in the dark (**Fig. 2a**,**c**). To reduce the background activity of paCas9-1, we replaced nMagHigh1-C714 with nMag-C714 because the combination of pMag and nMag shows lower background activity than that of pMag and nMagHigh1 (ref. 18). We transfected HEK293T cells with N713-pMag, nMag-C714 (named paCas9-2) and an sgRNA targeting *VEGFA* locus, and measured induced indel mutations in the light and the dark (**Fig. 3a**). The frequency of indel mutations in the dark induced by paCas9-2 was reduced to an undetectable level, as monitored using the T7E1 assay. Note that the light-induced indel frequency with paCas9-2 is comparable to that obtained using paCas9-1 (**Fig. 3b**). Therefore, we used paCas9-2 in all subsequent experiments.

Next, we tested whether paCas9 can enable light-induced spatial activation of genome editing (**Fig. 3c**,**d** and **Supplementary Fig. 7**). To visualize Cas9-induced NHEJ-mediated indel mutations in living cells, we used a surrogate EGFP reporter system that expresses EGFP fluorescence when a double-strand break is introduced into the target sequence by Cas9 (refs. 23,24). HEK293T cells transfected with paCas9-2, surrogate EGFP reporter and sgRNA targeting reporter were irradiated with slit-patterned blue light. After 24 h, a slit pattern of EGFP-expression was observed, showing that paCas9-2 can spatially control gene editing in response to light.

We also investigated whether paCas9 activation is reversible (**Fig. 3e**,**f**). We first transfected HEK293T cells with paCas9-2 and an sgRNA targeting *VEGFA*, and incubated these samples with blue-light irradiation to induce paCas9-2 activation. After 6 h, we split the cells and incubated half the cells in the dark and half the cells in light. After 6 h incubation in the dark or light, we performed a second transfection of an sgRNA targeting *EMX1*. The cells that had been incubated in the dark just before the second transfection were incubated again in the dark state, whereas the cells incubated in light before the second transfection were incubated again in light. After 30 h incubation, the genomic DNA was isolated and analyzed by T7E1 assay. Cells irradiated with blue light after the first transfection of paCas9-2 and with sgRNA targeting *VEGFA* showed indel mutations in the *VEGFA* locus, indicating that paCas9-2 was activated by the blue light. After the second transfection with an sgRNA targeting *EMX1*, cells irradiated continuously with blue light showed indel mutations in the *EMX1* locus; however, cells shifted to the dark did not have indel mutations in *EMX1* locus. This result indicates that paCas9-2 can be reversibly activated by blue light.

Several reports have shown that RNA-guided targeting of catalytically inactive Cas9 (dCas9) to a specific gene can sterically block RNA polymerase and transcript elongation, enabling RNA-guided gene silencing, which has been named CRISPR interference (CRISPRi)[25]. To further demonstrate the utility of paCas9, we tested photoactivatable and reversible control of CRISPRi. We named this method photoactivatable CRISPRi (paCRISPRi). To do this, we generated paCas9-2 containing D10A and H840A mutations (padCas9) (**Fig. 4a**). We also designed three sgRNAs targeting different regions of the CMV promoter-driven luciferase reporter containing PEST (proline-glutamate-serine-threonine rich) and mRNA destabilizing sequences[26]. padCas9 with each sgRNA targeting the luciferase reporter showed light-induced repression of luciferase reporter activity (**Fig. 4b**), demonstrating that our paCas9 platform also provides optogenetic control of RNA-guided transcription. We also found that the repression efficiency of padCas9 is slightly lower than that of full-length dCas9 (**Fig. 4c**). This is consistent with the fact that the frequency of indel mutation induced by paCas9 is ~60% of that achievable with full-length Cas9 (**Fig. 2a**). Efficiency might be improved by further engineering of padCas9 to, for instance,



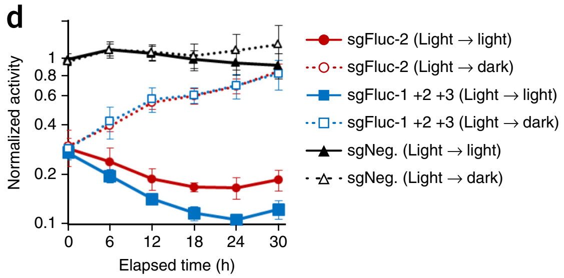


Figure 4 Optogenetic control of RNA-guided transcription interference with padCas9. (**a**) Schematic of paCRISPRi with paCas9-2 harboring D10A H840A mutations (padCas9). In the dark, N713 (D10A)-pMag and nMag-C714 (H840A) were inactive. Upon blue light illumination, pMag and nMag are heterodimerized, and, consequently, N713 (D10A) and C714 (H840A) are reconstituted as functional dCas9, enabling sgRNA-guided transcription interference. Gray, inactive; green, active. (**b**) padCas9 can repress gene expression in a light-dependent manner. HEK293T cells were transfected with N713 (D10A)-pMag, nMag-C714 (H840A), luciferase reporter and either indicated sgRNAs targeting luciferase (sgFluc-1, sgFluc-2 and sgFluc-3) or an unmatched control sgRNA (sgNeg.). 20 h after transfection, samples were illuminated by 1.2 W/m^2 blue light or kept in the dark for 30 h before measuring luciferase bioluminescence. In this experiment, a luciferase reporter containing PEST and mRNA destabilized sequence was used. Data are represented as the mean ± s.e.m. (n = 6 from two individual experiments with biological triplicates). Two-sided Welch's t-test was carried out. N.S., not significant; *** $P < 0.005$ versus the sample in the dark. (**c**) Comparison of the repression efficiency of padCas9 and full-length dCas9. Values are normalized to negative control with an unmatched control sgRNA (sgNeg.). Data are represented as the mean ± s.e.m. (n = 6 from two individual experiments with biological triplicates). Two-sided Welch's t-test was performed. N.S., not significant; * $P < 0.05$; *** $P < 0.005$ versus the full-length dCas9 in the dark. (**d**) Time-course of restored luciferase reporter activity after blue light irradiation. HEK293T cells were transfected with N713 (D10A)-pMag, nMag-C714 (H840A), luciferase reporter and the indicated sgRNAs. Immediately after transfection, samples were illuminated by 1.2 W/m^2 blue light for 30 h before measuring bioluminescence at time 0. After time 0 measurement, samples were illuminated by 1.2 W/m^2 blue light (solid lines) or kept in the dark (dotted lines), and bioluminescence was measured every 6 h. Data are shown as mean ± s.e.m. (n = 6 from two individual experiments with biological triplicates) normalized to negative control cells (under continuous light irradiation) at time 0.

optimize the protein stability, nuclear localization and dynamic range of photoinducible domains. We also tested whether padCas9-mediated gene repression could be switched off by halting light irradiation (**Fig. 4d**). We found that the reporter activity recovered gradually after we turned off the blue light, showing that CRISPRi with padCas9 is reversible. Collectively, our results show that paCas9 can offer spatiotemporal control of RNA-guided genome editing and transcription regulation in a reversible manner.

A rapamycin-inducible Cas9 designed with a split-Cas9 architecture that is similar to ours has recently been reported[6]. However, rapamycin-inducible Cas9 does not offer spatial and reversible activation because rapamycin diffuses freely and is hard to remove[27]. Furthermore, rapamycin can perturb endogenous mTOR signaling pathways[10].

The spatiotemporal and reversible properties of paCas9 are well suited for the dissection of causal gene function in diverse biological processes and for medical applications, such as *in vivo* and *ex vivo* gene therapies. Also, paCas9 has the potential to reduce off-target indel frequencies in Cas9-based genome editing. There are several studies showing that transient introduction of a Cas9:sgRNA complex prepared *in vitro* can improve the specificity of genome editing[7–9]. Because paCas9 can be switched off by stopping light irradiation, optically controlling the duration of paCas9 activation would contribute to reducing off-target gene modification.

This paCas9 platform has the potential to further facilitate CRISPR-Cas9 applications. For example, Cas9 applications *in vivo* have been limited owing to the packaging size limitation of viral vectors. Because the cDNA of our paCas9 components is shorter than that of full-length Cas9, it will be possible to separately package each paCas9 fragment into viral vectors having size limitation, expanding the opportunities of *in vivo* genome editing. Another potential application of paCas9 is intersectional control of Cas9 activity, enabling conditional genome editing with high precision. Recently, a conditional gene knockout strategy has been developed, by combining Cas9 and tissue-specific promoter[28]. By expressing each paCas9 component from two different tissue-specific promoters, Cas9 activity could be controlled by two promoter activities and light, enabling gene modification with ultrahigh precision.

We and other groups have recently reported dCas9-based photoactivatable transcription systems[29,30]. These dCas9-based optogenetic systems can activate targeted endogenous gene expression. Unlike these systems, the paCas9 described in this manuscript enables light-inducible targeted genome editing by NHEJ and HDR pathways. In addition, we show that paCRISPRi enables optogenetic control of targeted gene silencing.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We would like to thank J. Galipon for technical assistance. This work was supported by Platform for Dynamic Approaches to Living System from the Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT) and by grants from Japan Society for the Promotion of Science (JSPS).

AUTHOR CONTRIBUTIONS

Y.N. and M.S. conceived the project. Y.N., F.K., T.N. and M.S. designed the experiments. Y.N. performed experiments and analyzed data. Y.N. and M.S. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
2. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
3. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471 (2013).
4. Dow, L.E. *et al.* Inducible in vivo genome editing with CRISPR-Cas9. *Nat. Biotechnol.* **33**, 390–394 (2015).
5. González, F. *et al.* An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* **15**, 215–226 (2014).
6. Zetsche, B., Volz, S.E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33**, 139–142 (2015).
7. Kim, S., Kim, D., Cho, S.W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012–1019 (2014).
8. Zuris, J.A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing *in vitro* and *in vivo*. *Nat. Biotechnol.* **33**, 73–80 (2015).
9. Ramakrishna, S. *et al.* Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Res.* **24**, 1020–1027 (2014).
10. Laplante, M. & Sabatini, D.M. mTOR signaling in growth control and disease. *Cell* **149**, 274–293 (2012).
11. Brieke, C., Rohrbach, F., Gottschalk, A., Mayer, G. & Heckel, A. Light-controlled tools. *Angew. Chem. Int. Edn Engl.* **51**, 8446–8476 (2012).
12. Lee, H., Larson, D. & Lawrence, D. Illuminating the chemistry of life: design, synthesis, and applications of "caged" and related photoresponsive compounds. *ACS Chem. Biol.* **4**, 409–427 (2009).
13. DeRose, R., Miyamoto, T. & Inoue, T. Manipulating signaling at will: chemically-inducible dimerization (CID) techniques resolve problems in cell biology. *Pflugers Arch.* **465**, 409–417 (2013).
14. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
15. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
16. Kennedy, M.J. *et al.* Rapid blue-light–mediated induction of protein interactions in living cells. *Nat. Methods* **7**, 973–975 (2010).
17. Bugaj, L.J., Choksi, A.T., Mesuda, C.K., Kane, R.S. & Schaffer, D.V. Optogenetic protein clustering and signaling activation in mammalian cells. *Nat. Methods* **10**, 249–252 (2013).
18. Kawano, F., Suzuki, H., Furuya, A. & Sato, M. Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nat. Commun.* **6**, 6256 (2015).
19. Hsu, P.D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
20. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).
21. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).
22. Ran, F.A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
23. Kim, H. *et al.* Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat. Methods* **8**, 941–943 (2011).
24. Ramakrishna, S. *et al.* Surrogate reporter-based enrichment of cells containing RNA-guided Cas9 nuclease-induced mutations. *Nat. Commun.* **5**, 3378 (2014).
25. Qi, L.S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
26. Voon, D.C. *et al.* Use of mRNA- and protein-destabilizing elements to develop a highly responsive reporter system. *Nucleic Acids Res.* **33**, e27 (2005).
27. Lin, Y. *et al.* Rapidly reversible manipulation of molecular activity with dual chemical dimerizers. *Angew. Chem. Int. Edn Engl.* **52**, 6450–6454 (2013).
28. Shen, Z. *et al.* Conditional knockouts generated by engineered CRISPR-Cas9 endonuclease reveal the roles of coronin in *C. elegans* neural development. *Dev. Cell* **30**, 625–636 (2014).
29. Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H. & Sato, M. CRISPR-Cas9-based photoactivatable transcription system. *Chem. Biol.* **22**, 169–174 (2015).
30. Polstein, L.R. & Gersbach, C.A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nat. Chem. Biol.* **11**, 198–200 (2015).

ONLINE METHODS

Inducible Cas9 constructions. cDNAs encoding the N- and C-terminals fragments of codon-optimized *S. pyogenes* Cas9 fused with a nuclear localization signal from SV40 were amplified from Addgene plasmid 42230. cDNAs encoding FKBP and FRB were amplified from a human cDNA library. cDNA encoding CRY2PHR was amplified from Addgene plasmid 26871. The plasmid containing CIB1 was obtained from RIKEN BioResource Center (resource number: pda10875). cDNA encoding pMag, nMagHigh1 and nMag was prepared as previously described[18]. These inducible dimerization domains were amplified by standard PCR using primers that add glycine-serine linker sequences at the 5′ and 3′ ends. The inducible Cas9 constructs based on the N-terminal and C-terminal Cas9 fragments fused with dimerization domains were cloned into HindIII/EcoRI and HindIII/XhoI sites of pcDNA3.1 V5/His-A (Invitrogen), respectively. To construct paCas9 nickase and padCas9, we introduced D10A mutation in N-fragment of Cas9 and H840A in C-fragment of Cas9 by multisite-directed mutagenesis kit (MBL) according to manufacturer's directions. The full amino acid sequences of the paCas9-1 and paCas9-2 are shown in **Supplementary Note 1**.

sgRNA constructions. The sgRNAs targeting StopFluc reporters, *CCR5*, *EMX1*, *VEGFA AAVS1* and destabilized luciferase reporter were generated by annealed oligo cloning using BbsI site of Addgene plasmid 47108. The target sequences and oligonucleotides used for sgRNA construction are shown in **Supplementary Table 1**.

Reporter constructions. StopFluc reporters for plasmid HDR assay were constructed by inserting firefly luciferase sequence amplified from pGL4.31 vector (Promega) into HindIII and XhoI sites in pcDNA 3.1/V5-HisA and introducing stop codons and/or mutant PAM by the multisite-directed mutagenesis kit. The site-directed mutagenesis primers used to generate a series of StopFluc reporters are shown in **Supplementary Table 2**. The DNA sequences of StopFluc reporters are also shown in **Supplementary Note 2**. Luciferase donor vector was constructed by inserting inverted firefly luciferase sequence into XhoI and HindIII in bacteria-expression pColdI vector (Clontech). Destabilized luciferase reporter was constructed by inserting firefly luciferase with PEST sequence amplified from pGL4.31 vector into KpnI and XbaI sites in pcDNA 3.1/V5-HisA, and introducing 5 copies of mRNA-destabilizing nonamer sequence[26] (5′-TTATTTATT-3′) into XbaI and ApaI sites by annealed oligo cloning. Surrogate EGFP reporter was constructed by inserting mCherry and out-of frame EGFP into HindIII and XhoI sites in pcDNA 3.1/V5-HisA, and introducing *EMX1* target site between mCherry and EGFP using EcoRI and BamHI sites by annealed oligo cloning.

Cell culture. HEK293T and HeLa cells (ATCC) were cultured at 37 °C under 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM, Sigma-Aldrich) supplemented with 10% FBS (HyClone), 100 unit/ml penicillin and 100 μg/ml of streptomycin (GIBCO). Cell lines have not been tested routinely for *Mycoplasma* contamination.

Luciferase plasmid HDR assay. HEK293T cells were plated at approximately 2.0×10^4 cells/well in 96-well black-walled plate (Thermo Fisher Scientific), and cultured for 24 h at 37 °C in 5% CO_2. The cells were then transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocols. Plasmid encoding N-fragments of Cas9 fused with dimerization domain, C-fragments of Cas9 with dimerization domain, sgRNAs, StopFluc reporter and luciferase donor were transfected at a 2.5:2.5:5:1:4 ratio. The total amount of DNA was 0.2 μg/well. Twenty hours after the transfection, samples were incubated at 37 °C in 5% CO_2 under continuous blue light irradiation or in the dark. Blue light irradiation was performed using a 470 nm ± 20 nm LED light source (CCS Inc.). Intensity of blue light was 1.2 W/m^2. For chemically-inducible reassembly of split-Cas9, the culture medium was replaced with 100 μl of DMEM containing 10 nM of rapamycin instead of light irradiation. After 48 h incubation, the culture medium was replaced with 100 μl of phenol red-free DMEM (Sigma-Aldrich) containing 500 μM of D-luciferin (Wako Pure Chemical Industries) as a substrate. After 30 min incubation, bioluminescence measurements were performed using Centro XS^3 LB 960 plate-reading luminometer (Berthold Technologies). For comparison between DNA recognition ability of paCas9 and full-length Cas9, plasmids encoding full-length Cas9, sgRNAs, StopFluc reporter and luciferase donor were transfected at a 5:5:1:4 ratio. The total amount of DNA was 0.2 μg/well. After 48 h incubation, the culture medium was replaced with D-luciferin-containing phenol red-free DMEM and bioluminescence measurements were performed as described above.

Optogenetic genome editing experiments. For NHEJ-mediated indel mutation experiments, HEK293T cells were plated at approximately 1.0×10^5 cells/well in 24-well plate (Thermo Fisher Scientific), and cultured for 24 h at 37 °C in 5% CO_2. The cells were then transfected with Lipofectamine 2000 according to the manufacturer's protocols. Plasmids encoding N713-pMag, nMagHigh1-C714 and sgRNAs were transfected at a 1:1:1 ratio. As a positive control, plasmids encoding full-length Cas9 and sgRNAs were transfected at a 2:1 ratio. The total amount of DNA was 0.9 μg/well. Twenty hours after the transfection, samples were incubated at 37 °C in 5% CO_2 under continuous blue light irradiation or in the dark as described above. After 24 h incubation, genomic DNA was isolated using Blood Cultured Cell Genomic DNA Extraction Mini Kit (Favorgen) according to the manufacturer's instructions.

For HDR-mediated genome editing experiments, 6.0×10^5 HEK293T cells were nucleofected with 125 ng of N713-pMag, 125 ng of nMagHigh1-C714, 250 ng of sgRNA targeting *EMX1* and 10 μM of single-stranded oligonucleotide donor using the SF Cell line 4D-Nucleofector X Kit S (Lonza) and the CA-189 program. Transfected cells were plated at 2.0×10^5 cells/well in 24-well plate. Twenty hours after the nucleofection, samples were incubated at 37 °C in 5% CO_2 under continuous blue light irradiation or in the dark. After 48 h incubation, genomic DNA was isolated as described above.

In **Figure 3f** experiment, cells were plated and cultured as described above NHEJ-mediated indel mutation experiments. The cells were then transfected with Lipofectamine 3000 (Invitrogen) according to the manufacturer's protocols. Plasmids encoding N713-pMag, nMagHigh1-C714 and sgRNAs were transfected at a 1:1:1 ratio. The total amount of DNA was 0.5 μg/well. After 20 h incubation at 37 °C in 5% CO_2 in the dark, samples were incubated at 37 °C in 5% CO_2 under continuous blue light irradiation. After 6 h, we split and incubated the cells in dark and light state. After incubation for 6 h, we performed second transfection of sgRNA targeting *EMX1* with Lipofectamine 3000. The DNA amount was 0.5 μg/well. The samples in dark and light state just before the second transfection were incubated again in dark and light state, respectively. After incubation for 30 h, the genomic DNA was isolated as described above.

Mismatch-sensitive T7E1 assay for quantifying indel mutation of endogenous genes. The genomic region containing paCas9 target site was PCR-amplified using Pyrobest DNA polymerase (TaKaRa) using nested PCR for *CCR5* and *AAVS1* (First PCR: 98 °C, 3 min; (98 °C, 10 s; 55 °C, 30 s; 72 °C, 1 min) × 20 cycles; 72 °C, 3 min. Second PCR: 98 °C, 3 min; (98 °C, 10 s; 55 °C, 30 s; 72 °C, 1 min.) × 35 cycles; 72 °C, 3 min), two-step PCR with 5% DMSO for *EMX1* (98 °C, 3 min; (98 °C, 10 s; 72 °C, 30 s) × 35 cycles; 72 °C, 5 min) or touchdown PCR for *VEGFA* (98 °C, 3 min; (98 °C, 10 s; 72-62 °C, –1 °C/cycle, 30 s; 72 °C, 30 s) × 10 cycles; (98 °C, 10 s; 62 °C, 30 s; 72 °C, 30 s) × 25 cycles; 72 °C, 3 min). The primers for each gene are listed in **Supplementary Table 3**. The PCR amplicons were purified using FastGene Gel/PCR Extraction Kits (Nippon Genetics) following the manufacturer's protocol. Purified PCR products were mixed with 2 μl of 10 × M buffer for restriction enzyme (TaKaRa) and ultrapure water to a final volume of 20 μl, and re-annealed to form heteroduplex DNA (95 °C, 10 min; 90-15 °C, –1 °C/ 1 min). After re-annealing, heteroduplexed DNA were treated with 5 units of T7 endonuclease I (New England BioLabs) for 30 min at 37 °C and then analyzed by agarose gel electrophoresis. Gels were stained with GRR-500 (BIO CRAFT) and imaged with E-shot II gel imaging system (ATTO). Quantification was based on relative band intensities. Following equation is used to calculate the percentage of indel mutation by paCas9: $100 \times (1 - (1 - (b + c)/(a + b + c))^{1/2})$, where a is the intensity of the undigested PCR product, and b and c are the intensities of each T7E1-digested PCR product.

Sequence analysis. Purified PCR products used for the T7E1 assay were inserted into *Eco*RV sites in pcDNA3.1/V5-HisA vector. Plasmid DNAs were isolated by standard alkaline lysis miniprep, and sequenced using a T7 forward primer by the Sanger method.

RFLP assay for detecting HDR-mediated modification in endogenous human gene. The genomic PCR and purification were performed as described above. Purified PCR products were mixed with 30 units of HindIII (TaKaRa), 2 μl of 10 × M buffer for restriction enzyme and ultrapure water to a final volume of 20 μl, and incubated for 30 min at 37 °C. The digested products were analyzed by agarose gel electrophoresis. Gel staining and imaging were performed as described above. Quantification was based on relative band intensities. Following equation is used to calculate the percentage of HDR by paCas9: $100 \times (b + c)/(a + b + c)$, where a is the intensity of the undigested PCR product, and b and c are the intensities of each HindIII-digested product.

Spatial surrogate reporter activation. HEK293T cells were plated at 8.0×10^5 cells/dish on 35 mm dish (Iwaki Glass) coated with fibronectin (BD Biosciences), and cultured for 24 h at 37 °C in 5% CO_2. The cells were then transfected with Lipofectamine 2000 according to the manufacturer's protocols. Plasmids encoding N713-pMag, nMag-Cas9, sgRNAs targeting *EMX1* and NHEJ-mediated surrogate EGFP reporter containing *EMX1* targeting site were transfected at a 1:1:2:6 ratio. The total amount of DNA was 4.0 μg/dish. Twenty hours after the transfection, samples were illuminated by slit-patterned blue light using a photomask for 24 h at 37 °C in 5% CO_2. The width of slit is 2 mm. Cells were fixed with 4% paraformaldehyde in PBS for 15 min. Images were acquired using Axio Zoom.V16 stereo zoom microscope (Zeiss), and analyzed using Metamorph (Molecular Devices).

Photoactivatable CRISPR interference. HEK293T cells were plated at 2.0×10^4 cells/well in 96-well black-walled plate, and cultured for 24 h at 37 °C in 5% CO_2. The cells were then transfected with Lipofectamine 3000 according to the manufacturer's protocols. Plasmids encoding N713 (D10A)-pMag, nMag-Cas9 (H840A), mRNA-destabilized luciferase-PEST reporter and indicated sgRNAs targeting luciferase reporter were transfected at a 2.5:2.5:1:4 ratio. As a full-length dCas9 control, plasmids encoding full-length dCas9, mRNA-destabilized luciferase-PEST reporter and indicated sgRNAs were transfected at a 5:1:4 ratio. For the sample transfected with triple sgRNAs, the ratio of three sgRNAs was 1:1:1. The total amount of DNA was 0.1 μg/well. In **Figure 4b**,**c** experiments, 20 h after the transfection, samples were incubated at 37 °C in 5% CO_2 under continuous blue light irradiation or in the dark as described above. After 30 h, the culture medium was replaced with 100 μl of phenol red-free DMEM containing 500 μM of D-luciferin. After 1 h incubation, bioluminescence measurements were performed. In **Figure 4d** experiment, samples were incubated at 37 °C in 5% CO_2 under continuous blue light immediately after transfection. After 30 h, the culture medium was replaced with phenol red-free DMEM containing D-luciferin as described above. After 1 h incubation, light-illuminated samples were incubated again upon continuous blue light or in the dark, and bioluminescence measurements were performed at indicated time points.

Reproducibility and statistics. No sample size estimates were performed, and our sample sizes are consistent with that normally used in the genome editing and gene regulation experiments. No sample exclusion was carried out. No randomization was used. No blinding was used. In **Figures 3b** and **4b**,**c**, variances estimated by *F*-tests were equal, and two-sided Welch's *t*-tests were performed. We also confirmed the same statistical results by nonparametric two-sided Mann-Whitney tests because the normality of data was not tested. *P* values by Mann-Whitney tests are following; In **Figure 3b**, paCas9-1 and paCas9-2 in light showed no significant difference: $P = 0.773$. In **Figure 4b**, light and dark samples transfected with padCas9 targeting luciferase showed significant differences: $P = 0.00395$. Light and dark samples transfected with padCas9 and negative control sgRNA showed no significant difference: $P = 0.749$. In **Figure 4c**, padCas9 and full-length dCas9 transfected with sgFluc-1, sgFluc-2 and sgFluc-1+2+3 showed significant differences: $P = 0.00395$ (sgFluc-1), $P = 0.0065$ (sgFluc-2), $P = 0.00395$ (sgFluc-1+2+3), respectively. padCas9 and full-length dCas9 transfected with sgFluc-3 showed no significant difference: $P = 0.109$.

doi:10.1038/nbt.3245

SPONSOR FEATURE

E4TH EditForce

Produced with support from KISCO Ltd. in association with EditForce Inc.

The challenges faced by EditForce Inc., to go beyond genome editing

AUTHORS
Yusuke Yagi[1,2], Mitsuhiko Shirakawa[1], Takahiro Nakamura[1,2]

[1] EditForce Inc., 4th Fl. Tenjin Fukoku Seimei Bldg., Tenjin 1-9-17, Fukuoka 810-0001, Japan

[2] Faculty of Agriculture, Kyushu University, Fukuoka 812-8581, Japan

Address correspondence to:
EditForce Inc., 4th Fl. Tenjin Fukoku Seimei Bldg., Tenjin 1-9-17, Fukuoka 810-0001, Japan

http://www.editforce.jp

Yusuke Yagi: *yyagi@agr.kyushu-u.ac.jp*
Mitsuhiko Shirakawa: *mitsuh-s@editforce.jp*
Takahiro Nakamura: *tnaka@agr.kyushu-u.ac.jp*

EditForce Inc. provides an alternative option to genome editing and a novel tool for versatile editing of RNA molecules at the genomic scale, called 'transcriptome editing', based on pentatricopeptide repeat (PPR) protein engineering technologies. Our core technologies were invented at Kyushu University, Japan. The company was founded in May 2015 and is located in Fukuoka city in Kyushu, an island in south-west Japan. EditForce is an innovative company in the post-genomic era. Our mission is to provide novel DNA/RNA operating tools to understand and modify various living entities, for example plants and animals, and to translate our PPR technologies in various biological industries, including the pharmaceutical and agricultural industries.

New tools lead to a new world

Recently established genome editing technologies will open new avenues for biological research and development. EditForce Inc. was created to develop and apply DNA/RNA manipulation technologies using our core technology, PPR protein engineering, which includes PPR-based novel genome editing and versatile editing of RNA known as "transcriptome editing". Our technology enables sequence-specific manipulation of a single DNA molecule and a single RNA molecule in living cells. RNA manipulation could provide another layer of manipulation in various living entities, such as plants and animals. The core technology is based on PPR protein engineering, which was developed by Professor Nakamura's group at Kyushu University. Professor Nakamura is a founder of EditForce Inc. and has studied a plant-specific large family of PPR proteins. PPR proteins consist of a modular architecture of 35 amino acid repeat of PPR motifs, as observed in TALE proteins. Nakamura's group recently elucidated the principle behind the DNA/RNA binding mechanism and the recognition code[1], resulting in the development of new applications for engineered DNA and RNA binding proteins using DNA-binding and RNA-binding PPR proteins, respectively. EditForce Inc. was established in May 2015 through funding by KISCO Ltd (http://www.kisco-net.jp), which is a trading and business development company that is involved in advanced technologies related to chemicals, electronics, plastics and biotechnology, and most recently started a wide range of support for Spiber Inc. (http://www.spiber.jp). EditForce Inc.'s headquarters and the research facility are located in Fukuoka city in Kyushu in south-west Japan and the country's third-largest island (Figure 1). The goal of EditForce Inc. is to improve PPR-based technologies, to provide the research community and bioindustries with access to these technologies, in tight collaboration with Kyushu University, and to build a seamless pipeline from basic sciences to applied sciences. In May 2015, EditForce Inc started offering innovative molecular tools for DNA and RNA manipulations to understand and modify living entities and to translate PPR technologies for bioindustries including medicine, agriculture and bioproduction.

Figure 1 | EditForce Inc. is located in Fukuoka city, Japan. EditForce Inc. produces new genome/transcriptome editing tools based on pentatricopeptide repeat (PPR) proteins, which belong to a plant-specific RNA/DNA binding protein family.

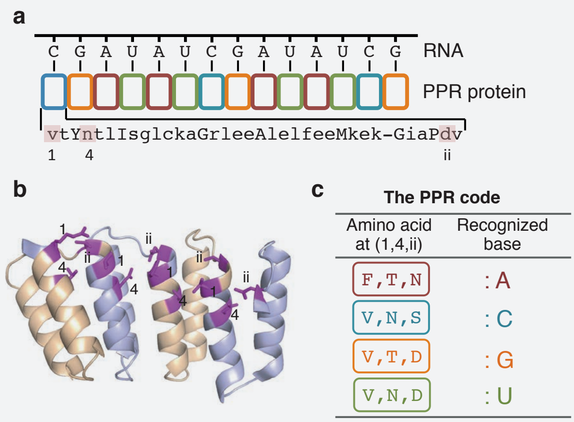


Figure 2 | PPR proteins. (a) Schematic representation of PPR–RNA interaction. A single PPR motif recognizes a single specific base that is defined by the combination of amino acids at three positions (1, 4, and ii, highlighted in red). **(b)** Tertiary structure of PPR array (Protein Data Bank; 4M59). **(c)** Examples of PPR codes for RNA binding.

Demands for new DNA/RNA manipulation tools in the post-genomic era

Modern molecular biology proposes a central dogma that explains the flow of genetic information from DNA to RNA to protein. The universality of the central dogma indicates that all truth and enigma of biological systems are hidden in the genome sequence. Strong interest in revealing genetic information has driven whole genome sequencing and advances in DNA sequencing technologies. The development of high-throughput sequencing methods, called next generation sequencing (NGS), resulted in the acquisition of information from large DNA sequences at once. For example, the human genome, which is composed of three billion base pairs, can be sequenced within one day using this technique. NGS provided a huge amount of genome sequence information for numerous living organisms, such as animals, fish, plants and insects. NGS has also been used to identify minor differences within genome sequences among individuals, which may be involved in susceptibility to disease and pathogens or sensitivity to chemicals. In our post-genomic era, suitable molecular tools are in demand to understand and utilise this vast amount of information. Recently established genome editing technologies are DNA manipulation tools that enable the modification of a single genetic locus within the genome in a living cell.

Despite the central dogma principle, important variations from DNA occur in RNA, especially in eukaryotic cells. This includes a series of processing steps, including RNA splicing, control of localisation, and translation efficiency that determine how much and when a protein is produced. Thus, RNA provides diverse and crucial functions allowing eukaryotic cells to read their genomes. Moreover, the genome is largely occupied by 'junk DNA' in the animal genome. For example, the proportion of protein coding genes represents only 1.5% of the human genome. However, simultaneous analysis of the transcriptome (gene expression studies) revealed that a large part (more than 60%) of the animal genome is transcribed into RNA. A huge amount and a variety of non-coding RNAs have been identified, which are transcribed, but not translated into proteins. The functions of most non-coding RNAs still remain unclear. Protein coding genes have been found to be more complex than originally thought. Through alternative splicing, various products can be produced from a single genomic locus. One challenge in biology during this century is to understand how DNA/RNA information is utilised by living cells and how we can effectively use this information. Our PPR-based RNA manipulation technique has promising potential to be a novel versatile RNA manipulation tool in response to this demand.

PPR proteins, a priority of EditForce's technology

PPR protein is the abbreviation for pentatricopeptide repeat (PPR) motif-containing protein. The 'Penta-Trico-Peptide' denotes 35 amino acids. A PPR protein typically consists of a repeating unit of 35 amino acids of PPR arrays (Figure 2). The PPR motif and PPR protein were first recognised in 2000, by informatics studies during the genome sequencing project of a model plant, *Arabidopsis thaliana*. The PPR protein genes were identified as a gene family that is only found in land plants (500 variations of PPR genes/plant)[2]. A series of analyses demonstrated that PPR proteins are involved in various gene expression events in a sequence-specific manner. Typically, a single PPR protein targets a single RNA or DNA molecule and is required for RNA stability, RNA splicing, processing, RNA editing, or transcription of chloroplast or mitochondrial genes[3]. In plants, most PPR proteins (90%) seem to be target RNA molecules, therefore functioning as RNA-binding PPR or canonical PPR proteins, but the remaining PPR proteins act as DNA-binding PPR proteins.

Our interest was to determine how PPR proteins act as sequence-specific RNA-binding proteins. PPR proteins contain tandem arrays of 2–30 PPR repeats that fold into a pair of anti-parallel α helices (Figure 2). The PPR motif array forms a super-helical binding surface similar to that observed in the transcription activator-like effector (TALE) protein. The principle of PPR–RNA recognition has been elucidated by focusing on the relationship between the motifs and binding sequences. One PPR motif corresponds to one nucleotide and the combination of amino acids at three particular positions (residues 1, 4, and ii) defines the nucleotide specificity in a programmable manner[1] (Figure 2b and 2c). This knowledge provides the rationale for the engineering of custom RNA binding proteins, which correspond to arbitrary RNA sequences.

Furthermore, we found that the PPR–RNA recognition principle could be applied to DNA-binding PPR proteins, by collaborating with Prof. Yamamoto's group at Hiroshima University. Thus, we have established engineering platforms for both custom DNA and RNA binding proteins. EditForce Inc.'s first challenge was to bring these new PPR-based tools outside of the plant organelles.

DNA-binding PPR as a new genome-editing tool

Genome modification techniques have been used by humans for millennia and have been beneficial. A good example is plant breeding in agriculture over the past 6,000 years. Rational DNA manipulation began with the discovery of restriction enzymes in the 1960s. Recently established genome editing technologies using zinc finger, TALE, or CRISPR have enabled rational modification of specific gene(s) in living cells[4,5] (Box 1). Our DNA-binding PPR-based technology is protein-based and can be used as TALE proteins, due to the PPR modular

structure with a definite DNA-recognition code in a motif context independent manner (Figure 3). PPR proteins present several advantages. The amino acid sequence of natural PPR proteins is less conserved and permits PCR-based amplification or chemical DNA synthesis and PPR proteins are free during the selection of the target DNA sequence, in contrast to the requirement for a 5′-T residue for TALE or the PAM sequence of CRISPR/Cas. Improving the activity of DNA-binding PPR and the useful construction method are the two major tasks to be tackled.

RNA-binding PPR as a novel versatile RNA manipulation tool for transcriptome editing

In contrast to the progress in DNA manipulation tools for genome editing, equivalent RNA manipulation tools that can operate on a single specific RNA in a living cell are still in their infancy. Available RNA manipulation tools can be divided into two classes, RNA-based or protein-based tools. Currently, the most prominent RNA manipulation tools are the guide RNA-based siRNAs used for RNA interference, which enables sequence-specific transcriptional or post-transcriptional gene silencing and is available in most eukaryotes by introducing double-stranded RNAs. This silencing technology has been widely applied in research. However, the therapeutic use of nucleic acid-based technologies, including RNAi, has been hampered due to the instability of RNAs *in vivo*, absence of suitable delivery systems, unwanted innate immune responses, and off-target side effects[6]. Moreover, these technologies are limited to 'knockdown' applications, and cannot be used to switch on the expression of an mRNA or to alter the function of an encoded gene product. Consequently, the development of alternative, protein-based technologies would be highly desirable. Several efforts to alter gene expression by targeting RNA with modified RNA binding proteins (RBPs) have been attempted by various scientists. However, the proteins were unable to operate on endogenous mRNAs because of the unavailability of RBPs with arbitrary specificity.

Pumilio/fem-3 (PUF) protein appeared suitable for the construction of designer RBP[7]. PUF proteins typically consist of eight tandem repeats of a 36 amino acid unit. The structure of PUF–RNA complexes presents a simple 'one repeat to one nucleotide' correspondence, with RNA-contact defined by two amino acid residues at specific positions. Several investigators

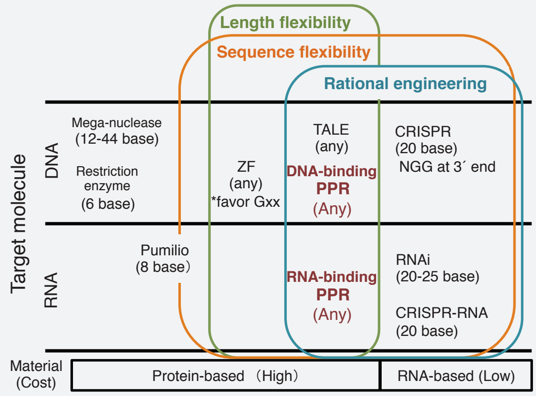


Figure 3 | Characteristics of available DNA/RNA manipulation tools. The custom DNA-binding PPR is similar to transcription activator-like effector (TALE). RNA-binding PPR is a protein-based, unique technology, enabling the engineering of arbitrary sequences of any length. The number in parentheses indicates the target sequence length.

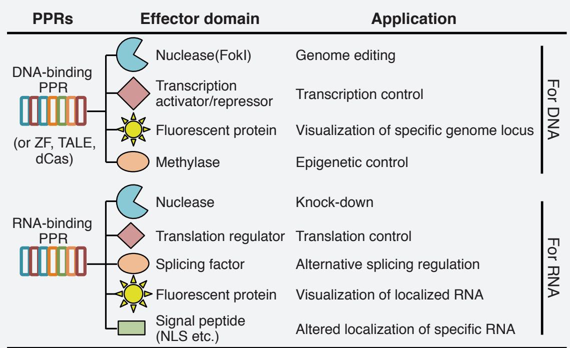


Figure 4 | Possible DNA/RNA manipulation tools based on PPR proteins. Various DNA/RNA manipulation tools can be designed by fusion of sequence-specific DNA/RNA binding modules and effector domains with various functions. EditForce Inc. provides a wide variety of DNA and RNA manipulation tools using PPR proteins. Zinc finger (ZF), TALE, or dCas (inactivated form of clustered regularly interspaced short palindromic repeat-associated (Cas)) are also applicable for DNA manipulation tool engineering.

successfully engineered PUF proteins to recognise new target sequences. Custom PUF repeat proteins have been conjugated with various effector domains for RNA cleavage, RNA visualisation, control of translation and alternative splicing. Therefore, these proteins are appealing as useful protein-based tools[7]. However, the flexibility of the binding sequence alteration was unexpectedly restricted due to the context dependency of PUF as observed for the DNA-binding zinc finger module. Additionally, PUF is responsible for only eight to nine nucleotide sequences, suggesting that its versatile use is significantly impaired.

The RNA-binding PPR motif has emerged as being exceptionally promising for the rational design of sequence-specific RBPs. PPR proteins bind nucleic acid with a 1:1 ratio between protein motifs and RNA bases, with a definite RNA recognition code, determining a couple of key sites in each motif[1]. Natural PPR proteins consist of 2–30 repetitions of PPR motifs, suggesting that they might be applicable to target sequences of various length[8] (Figure 3). Moreover, their huge natural diversity in plants provides plenty of resources for engineering. The PPR-based versatile RNA manipulation tool could provide a novel transcriptome editing technique via the design of new RNA manipulation tools.

Potential of DNA/RNA manipulation tools

Recently designed genome-editing tools using zinc finger, TALE and CRISPR, which enable the rational deletion or insertion of a DNA sequence *in vivo*, open a new avenue in biology. Although genome editing mainly uses the

Box 1 | The history of gene modification techniques

Plant breeding is an ancient gene modification technique that is performed by crossing two strains with different genetic backgrounds to produce new crops and vegetables with desirable characteristics such as high and stable yield, good taste or beautiful petal colours. The breeding processes still remain time- and labour-consuming as these processes largely rely on the occurrence of random DNA alterations.

Restriction enzymes are protein-based DNA scissors, discovered in the 1960s. However, their short recognition sequence (typically a 6 base pair recognition site) has restricted their application at the genomic scale. The question remains: What DNA sequence length is required to select a single specific locus from the human genome, which is three billion base pairs long? This can be theoretically estimated to be 16 base pairs, because DNA presents four variations of A, T, C, and G ($1/4^{16}$, approximately one site in four billion base pairs). In an attempt to solve this problem, several approaches were developed using a meganuclease that cleaves a 14–44 base pair-specific DNA sequence. However, alteration of the sequence specificity is severely restricted.

The first DNA scissors with arbitrary sequence recognition technology was the zinc finger-based technology[4]. A single zinc finger module recognises three nucleotides, and the repeating architecture of the zinc finger protein can be used to combine the zinc finger modules to obtain specific sequence recognition with a comfortable length, for example the 6 repetitions correspond to an 18-nucleotide DNA element, recognizing a single genomic locus within the genome. However, the zinc finger module originally prefers the 'GNN' sequence and its specificity is altered by the adjoining zinc finger module (motif-context dependency). Therefore, it was difficult to design and construct a zinc finger presenting the intended DNA-binding capacity (Figure 3).

These limitations have been largely solved by TALE-based technologies[4]. TALE presents a modular structure of 34 amino acid unit repeats. A single TALE unit corresponds to a single nucleotide and the two positions of amino acids determine the recognizing DNA base in a programmable and motif-context-independent manner. Since TALE enabled the rational design of the molecular scissors, genome-editing technology was applied to various organisms, including animals, plants, fish and bacteria. Limitations of the TALE technology include the use of a peculiar cloning method of Golden Gate assembly due to the high amino acid conservation and the requirement of a thymine (T) residue at the 5′-end of the target DNA sequence, due to the presence of the un-exchangeable N-terminus region.

CRSIPR/Cas (clustered regularly interspaced short palindromic repeat/CRISPR associated) then drastically improved genome-editing techniques[5]. This technique uses a guide-RNA based system that enables fast and cost-effective genome editing by just designing an oligonucleotide of 20 nucleotides that is complementary to the target DNA sequence. This system is also applicable for multiple gene modifications at once. Its simplicity and power overcome the limitation that a PAM sequence (-NGG) at the 3′-end of the target sequence is required and suspicions regarding its high off-targeting rate.

fusion of the DNA-binding module to a DNA scissor (nuclease), the potential of DNA-binding modules is huge, because these DNA tools can be easily modified to design different DNA manipulation tools by exchanging the nuclease domain with other functional domains such as transcription repressor/activator, epigenetic regulator, or fluorescent protein. Equally with DNA manipulation tools, the PPR-based RNA binding module can be connected to various effector domains including ribonuclease, translational activation or repression domain, localisation domain, or fluorescent protein to visualise endogenous RNA (Figure 4). Thus, PPR-based RNA manipulation tools provide novel opportunities to control endogenous RNAs. These DNA/RNA manipulation tools should considerably facilitate genome structure and function studies and related industrial applications. Moreover, combining the use of multiple DNA/RNA manipulation tools would help to create more elaborate and effective controls of the biological process of interest.

EditForce vision of PPR engineering

EditForce focuses on research and development of molecular tools operating both on DNA and RNA, based on PPR technology (Figure 5). The potential would drive innovations in various biological research fields and industries.

For basic sciences: The human genome sequence was determined in 2003. The number of protein-coding genes is approximately 25,000, occupying 1.5% of the total genome sequence. The number is unexpectedly low and almost equivalent to that in mouse and plants, indicating that the number of protein-coding genes does not explain the complexity and identity of living organisms. A parallel comprehensive study of the transcriptome discovered an unimagined complex, novel feature of the new RNA continent, including a huge amount of non-coding RNAs and extensive alternative RNA splicing, especially in mammals. Therefore, RNA is an important key factor to understand the complexity and identity of human beings. Editing tools for DNA and RNA can be applied to investigate the genome complexity and the new RNA continent.

Furthermore, an emerging research area in synthetic biology has received special attention. Synthetic biology is an interdisciplinary area developed to design and construct biological devices or biological systems. Together with the recent advances in the chemical synthesis of long DNA sequences, the field of 'genome design', involving more drastic modifications of the genome sequence or the chemical build-up of the whole genome sequence, could be a realistic technology. Genome editing tools, including zinc finger, TALE, CRISPR/Cas, and DNA-binding-PPR, facilitate the design of genome and transcription circuits that initiate gene expression. Furthermore, our PPR-based RNA manipulation tools would add operation systems at the RNA level. Combining DNA and RNA operation tools will lead to a new technology for 'genome programming'. Various DNA/RNA manipulation tools will open the door of genome design/programming beyond genome editing.

For medicine: Genome editing has been anticipated in the medical field, especially for reproduction and gene therapy. Several attempts have already started using TALE and CRISPR, including chimeric antigen receptor (CAR) T cell immunotherapy and the generation of disease-relevant cells using pluripotent stem cells. We are also focusing on these relevant areas. Moreover, our protein-based RNA manipulation tools using RNA-binding PPR can be applied to acquired diseases, which many persons may be suffering from, unlike genetic diseases. The discovery of RNAi resulted in a strong driving force, establishing RNA-targeted pharmacology. However, siRNA-based pharmacology has been hampered by the difficulties described earlier (off-target effects, absence of the appropriate delivery systems,

and unstable traits due to the chemical nature of RNA). The PPR-based technology is protein-based and could therefore overcome many disadvantages of RNA-based tools. EditForce Inc. has established a translation enhancement tool for specific RNA, which may be applied to reinforce the activity of the tumour suppressor gene, *p53*, for cancer therapeutic purposes, for example. The development of various RNA manipulation tools for up- and down-regulation, splicing, and editing (RNA modification) are in progress. These tools can be utilised for novel therapeutic use, not merely for the suppression of gene expression as for siRNA, but also for the up-regulation or change in the coding capacity without affecting the patient's genome. Moreover, we succeeded in designing a custom RNA-binding PPR molecule that specifically recognises a human infectious RNA virus. This RNA-binding PPR molecule will be used for detection and therapy in the near future.

For agriculture: genome editing can also be used in agriculture. In fact, genome editing is listed as one of the new plant breeding techniques, enabling fast and precise genetic modifications without introducing foreign genes in the final product[9]. Therefore, genome-editing crops could be out of the current legal regulation based on the Cartagena protocol on Biosafety. Although regulation of genome-editing plants is currently under debate, PPR-based genome editing is also applicable to agricultural breeding. PPR research and development for agricultural purposes has been in part supported by the Cross-ministerial Strategic Innovation Promotion Program (SIP) of the Council for Science, Technology and Innovation (CSTI) in Japan. Furthermore, we have developed a F1 hybrid seed product using our RNA manipulation technology. The F1 hybrid seed production aimed at adding hybrid vigour, also called heterosis, in various cultivars. In the F1 hybrid seed production, cytoplasmic male sterility (CMS) is used as an important agricultural trait[10]. The CMS trait is caused by the expression of an aberrant protein-coding gene in the mitochondrial genome, but the presence of a nuclear gene of restorer-of-fertility (Rf) cancels the sterility by post-transcriptional repression of the aberrant mitochondrial gene. The CMS-Rf system has been used in various cultivars, including those of maize, canola, and rice, for a long time. Various Rf genes have been identified and, in most cases, assigned as RNA-binding PPR protein genes. Elucidation of the PPR-RNA recognition mechanism led to the design of artificial Rf genes that enable more efficient use of the CMS-Rf system or the application of the CMS-Rf system to crops in which this system has not yet been applied. Mitochondrial genome editing is also one of our research and development interests for the purpose of designing artificial CMS genes and developing pairs of artificial Rf and CMS genes for a wide range of crops.

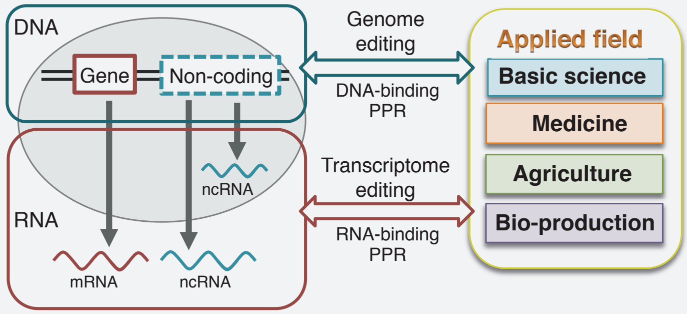


Figure 5 | EditForce Inc. vision is to use PPR-based DNA and RNA manipulation technologies to edit the genome and transcriptome in living cells. We provide genome and transcriptome editing tools for various fields of biological sciences, including basic science, medicine, agriculture, and bio-production.

For bio-production: The production of various materials, special proteins or chemical compounds for a wide range of applications such as biofuels and therapeutic uses has been utilised in living organisms, including bacteria, cultured cells and plants. Homologous systems in which the compound of interest is uniquely produced by the organism itself or by heterologous expression systems in which foreign protein(s) are expressed using the appropriate host organism as a bioreactor are utilised. Genome editing technologies enable the application of these technologies to various organisms and should facilitate the engineering of host organisms to improve bio-production. EditForce Inc.'s technologies, editing of both DNA and RNA, enable the optimisation of the genome (DNA) and the expression circuit (DNA and RNA). The PPR-based translational activation tools, described above, have been used for the production of a useful protein of therapeutic interest. The development of another molecular tool operating in subcellular localisation, and nuclear-cytoplasmic transport, is ongoing. These RNA operation tools would open other opportunities for cell engineering in bio-production.

In conclusion, the development of genome editing tools should allow us to better understand and effectively use the huge amount of genetic information currently available. EditForce provides another option for genome editing tools and a new technology for transcriptome editing. Combining the use of DNA and RNA manipulation tools will open a new field of genome design/programming. EditForce Inc. provides genome design/programming tools, which will lead to a better understanding of biology and can be applied in a wide range of industries.

REFERENCES

1. Yagi, Y. *et al*. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One* **8**, e57286 (2013).
2. Small, I.D. & Peeters, N. The PPR motif – a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci.* **25**, 46–47 (2000).
3. Nakamura, T. *et al*. Mechanistic insight into pentatricopeptide repeat proteins as sequence-specific RNA-binding proteins for organellar RNAs in plants. *Plant Cell Physiol.* **53**, 1171–1179 (2012).
4. Perez-Pinera, P. *et al*. Advances in targeted genome editing. *Curr Op Chem Biol.* **16**, 268–277 (2012).
5. Doudna, J.A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
6. Davidson, B.L. & McCray, P.B. Jr. Current prospects for RNA interference-based therapies. *Nature Rev Genet.* **12**, 329–340 (2011).
7. Wang, Y. *et al*. Engineered proteins with Pumilio/fem-3 mRNA binding factor scaffold to manipulate RNA metabolism. *FEBS J.* **280**, 3755–3767 (2013).
8. Yagi, Y. *et al*. The potential for manipulating RNA with pentatricopeptide repeat proteins. *Plant J.* **78**, 772–82 (2014).
9. Lusser, M. & Cerezo, E.M. Comparative regulatory approaches for new plant breeding techniques, *JRC Scientific and Technical Reports*. (http://ftp.jrc.es/EURdoc/JRC68986.pdf) (2012).
10. Chase, C.D. Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. *Trends Genet.* **23**, 81–90 (2007).

3 December 2015
nature.com/diagnostics-modelling

Infectious disease control and elimination
Modelling the impact of improved diagnostics

BILL & MELINDA GATES foundation

DIAGNOSTICS MODELLING • CONSORTIUM

npg nature publishing group

INTRODUCTION OPEN

Expanding the role of diagnostic and prognostic tools for infectious diseases in resource-poor settings

Azra C. Ghani[1], Deborah Hay Burgess[2], Alison Reynolds[1] & Christine Rousseau[3]

Nature 528, S50–S52 (3 December 2015), DOI: 10.1038/nature16038
This article has not been written or reviewed by *Nature* editors. *Nature* accepts no responsibility for the accuracy of the information provided.

The life-saving impact of new diagnostic and prognostic technologies that aim to reduce the burden of infectious diseases is often not well understood. Although the potential benefits of other interventions such as drugs and vaccines can be estimated by simply counting the numbers treated and multiplying that by the effect size of the intervention, understanding the role that diagnostics can have requires more complex analyses. As for other interventions, the performance of the tool is important. Few, if any, diagnostic tools have 100% sensitivity and specificity or a perfect quantitative range. However, unlike drugs or vaccines, the impact of the diagnostic depends on the actions taken after the diagnostic or prognostic test result. First, the tests may not always be run or interpreted correctly because they are often used by staff with minimal training. This may further reduce performance, depending on the level of the health-care setting in which the test is used. This has been clearly demonstrated by a South African study[1] in which several HIV rapid diagnostic test procedures were observed and only 3.4% were found to have been performed in full compliance with procedure, suggesting that there is a potential for high rates of misdiagnoses. Second, the clinician or health-care worker must interpret the results and make the appropriate clinical decision. In the case of Cepheid Gene Xpert and malaria rapid diagnostic tests, for example, studies have shown that even when the test gives the correct result, treatment is often provided empirically[2,3]. Third, the clinical decision needs to be realized. This will depend on the availability of appropriate treatment facilities and drug stocks. Crucially, the combination and timing of these processes can affect the onward transmission of infectious diseases at the population level and hence have an impact on the control of epidemics or progress towards elimination of endemic diseases.

This complexity adds to the controversy in assessing the value of diagnostics and often delays the already long process of discovery, development and delivery of new technologies for global infectious diseases. This was addressed in the 2006 Nature Publishing Group supplement *Improved Diagnostic Technologies for the Developing World*, which used modelling techniques to define the value of new diagnostic tools for resource poor settings[4]. Over the subsequent 10 years there has been encouraging progress in the development and use of new diagnostics, but many gaps remain. By way of a response, this collection presents new modelling work that addresses the potential impact of diagnostic tools both at the individual and population level.

This work could not have come at a more crucial time. Over the past decade there has been a shift in the epidemiology of infectious diseases, with dramatic reductions in burden, which was catalysed by the Millennium Development Goals and the associated increase in global health funding[5]. This has been accompanied by a shift from control of diseases in centralized health-care settings to prevention and early treatment. Accompanying this changing epidemiology, diagnostics are increasingly demanded and used in novel health-care paradigms. Technological advances have supported this shift. For example, the reach of centralized laboratory testing can be extended through the use of specimen collection and stabilization technologies in combination with sample transport systems; such as the use of dried blood spots for HIV levels, malaria parasite detection and serology[6–8]. In addition, new portable and integrated technologies can allow testing at primary health facilities, providing greater access to care and adoption[9]. This was first quantified in a seminal study from Mozambique where point-of-care CD4 technologies with same-day results enabled a near doubling of patients on treatment, owing to the reduction in loss to follow-up that normally occurs when patients wait several weeks for results from centralized laboratory testing[10]. For chronic conditions such as HIV, self-testing is also being considered as a method to improve testing coverage and ultimately linkage to care[11]. Rapid results and ease of use are clearly key characteristics that are affected by the treatment paradigm or patient flow. For example, the balance may be shifted towards sensitivity over specificity if the reason for testing is to determine onward referral rather than immediate treatment. For reasons such as these, the target product specification for a diagnostic in these settings is likely to differ to that in centralized health-care settings.

As reported in the 2006 supplement, modelling can be a useful tool to capture the health impact that improved diagnostics can have on global health efforts. However, although the decision-analytic approach previously adopted was appropriate to estimate the impact of a diagnostic at the point of care on health outcomes such as cases, deaths and disability-adjusted life years (DALYs), in the wider contexts it is also important to capture the effect of the diagnostic on onward transmission. Transmission dynamic models are well developed for such a purpose and are increasingly used to guide product development and public-health decision-making for a wide range of diseases. However, the integration of diagnostics into such models is often overlooked. To fill this gap, the Diagnostics Modelling Consortium funded by the Bill and Melinda Gates Foundation was formed in 2013 to catalyse the incorporation of diagnostics into transmission dynamic models across key global health diseases, including HIV, tuberculosis, pneumonia, malaria and neglected tropical diseases. The Consortium and its partners brought together not only modelling groups, but also those involved in diagnostics development and disease

[1] MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, Norfolk Place, London W2 1PG, UK. [2]Principal, Kimataifa Diagnostics and Devices Consulting, 17319 15th Drive SE, Mill Creek, Washington 98012. Seattle, USA. [3]Bill and Melinda Gates Foundation, 500 5th Avenue North, Seattle, Washington 98109, USA. Correspondence should be addressed to A. C. G. e-mail: a.ghani@imperial.ac.uk.

Table 1 | Levels of laboratory testing available for public health programmes in different levels of the health system. Adapted from ref. 12.

Health-care level	Description	Appropriate diagnostic or prognostic tools
0	Informal – 'under the tree'	• First point of care with a community health worker – tool must be simple to use and not require special storage • Prognostic tools particularly relevant for rapid referral
1	Primary – health post and centres	• Simple diagnostic techniques, including collection of dried blood spots and rapid or dipstick tests
2	District – district hospital	• Act as referral centre for specimens sent from level 1 • Include dedicated laboratory space, trained technicians and reagents • Can manage a more extensive test menu for diagnosis and treatment
3	Regional or provincial – referral hospitals or part of regional or provincial health bureau	• Laboratory facilities sufficient to perform complete menu testing for HIV/AIDs, tuberculosis and malaria as well as many other diseases • Typically include level 2 laboratories
4	National or multicountry – reference laboratories for one or more countries	• Strengthen laboratory capacity for all diseases of concern and provide molecular and esoteric testing that cannot be performed in level 2 laboratories, for example nucleic acid assays, HIV drug resistance studies and tuberculosis drug susceptibility studies

specialists who could define the strategic needs within these priority disease areas. Over an 18-month period the groups worked together to define the questions that, when answered, would best inform diagnostic product and prognostic tool development, and to extend existing models to address these questions. At the same time, the group sought to share experiences and lessons learned across the disease areas.

Two themes emerged in the subsequent work. The first was the importance of considering the patient flow for use of both diagnostic and prognostic tools in the wider community. The papers by Floyd *et al.* and Arinaminpathy and Dowdy describe the importance of capturing individuals in the wider community who do not promptly seek care. Floyd *et al.* assess the potential impact of a new prognostic device — pulse oximetry — for pneumonia, and find that simple medical devices that increase early prognosis of severe pneumonia could have a substantial public health impact as well as being highly cost-effective, provided subsequent access to oxygen treatment is available. Arinaminpathy and Dowdy also consider this issue, but more broadly, for tuberculosis, arguing that the evaluation of new diagnostics needs to take into account not only the sensitivity and specificity of the diagnostic itself, but also the impact that it can have on patient behaviour and care seeking.

Similarly, the Working Group on Modelling of Antiretroviral Therapy Monitoring Strategies in sub-Saharan Africa explore the potential public health impact and cost-effectiveness of using viral load measurement to differentiate levels of care so that those with a lesser need visit the clinic less often, thereby freeing health-care capacity for those in greater need. Despite the limitations associated with viral load testing using dried blood spots (and noting that point-of-care tests may become available in the future), they find that such an approach is cost-effective. In the second HIV article, Sharma *et al.* present a systematic review of the methods that have been used to improve coverage of HIV testing. They find that compared with facility-based testing, community testing and counselling is a model that identifies HIV-infected individuals at an earlier stage of infection (higher CD4 counts). In addition, they find that mobile and self-testing are even more effective in reaching key population groups, including men, young people and those at higher risk.

Efforts have been made to standardize the settings where diagnostic technologies can be used, given the range of levels that exist — from centralized laboratories to minimally-resourced settings. During a meeting in January 2008 held in Maputo, Mozambique, the World Health Organization brought together key stakeholders who were charged with making recommendations on laboratory standardization and harmonization[12]. This group defined four tiers of the laboratory system (see Table 1), as well as level 0 or 'under the tree' — an informal site where diagnostics can and should be used. As patients flow through these levels, a range of different technologies will probably be used to meet their needs, and it will therefore be important to optimize the tools' placement and use based on potential impact. Now that technologies are available that can integrate into each laboratory setting, next-generation modelling efforts will need to address optimal placement.

The second theme was the use of new diagnostic technologies to target interventions with the purpose of disease elimination. This comes at a time when new global commitments for the elimination of diseases have been made, including a call in 2007 by Bill and Melinda Gates to move towards malaria eradication[13] (and subsequent inclusion of elimination goals in the World Health Organization Global Technical Strategy[14]) alongside the elimination goals set out for neglected tropical diseases under the London Declaration[15]. Diagnostic needs for elimination pose new challenges. First, as the disease declines to low levels, identifying remaining foci of infection is important. For this, diagnostic tools need to be sufficiently sensitive to detect the remaining reservoir of infection. However, as the end point is reached, and in the subsequent maintenance phase, identifying infected individuals becomes crucial. For this, highly sensitive and rapid diagnostic tests are required. In both the elimination and the maintenance phase, diagnostics have a crucial role in the overall surveillance strategy, providing the first indication of potential for re-emergence. However, for many diseases appropriate diagnostics for this phase are not yet developed, with major challenges remaining in relating the different biomarkers to disease status. For example, although substantial progress has been made in moving towards the elimination of onchocerciasis in West Africa, current diagnostic tools — such as skin-snips to detect micro-filariae and nodule palpation to identify foci of transmission — are unlikely to identify very low levels of infection and thus may have insufficient sensitivity to prevent resurgence[16]. Hence, for diseases such as this it is likely that a combination of diagnostic tests will need to be used, each targeted to the appropriate stage of transmission.

These issues are addressed in two related articles on *Plasmodium falciparum* malaria. Wu *et al.* undertake an extensive review of the relationship between current diagnostic tools used in endemic settings. They find that current rapid diagnostic tests detect only 41% of infections detected by high-sensitivity polymerase chain reaction (PCR) techniques, indicating that a substantial number of infections will be missed if this diagnostic is used in the field and implying that a large infectious reservoir remains. Slater *et al.* continue this theme to estimate the target product specifications for new diagnostic tests that aim to reduce onward transmission. They find that increasing the sensitivity of the current rapid diagnostic test tenfold could detect 83% of the infectious reservoir. Applying this strategy to settings in sub-Saharan Africa and Asia, the authors demonstrate that increase in sensitivity could widen the areas in which mass screen-and-treat programmes and targeted mass drug administration could succeed in interrupting transmission.

Medley *et al.* take a similar approach to assessing the role of diagnostics for visceral leishmaniasis — concentrating on the potential for elimination in the Indian subcontinent. They find that shortening the time from health seeking to diagnosis could dramatically reduce incidence, especially if a diagnostic can be developed that is able to detect infected individuals before the onset of clinical kala-azar. The study also highlights the importance of bringing modellers and scientists together to develop diagnostic tools early on in the development process. Given the overall poor understanding of the aetiology and transmission biology of pathogens such as *Leishmania*, modelling can help to identify key parameters for which further data are needed. In turn, the data collected can be used to refine the models in subsequent iterations, potentially speeding up the process of diagnostic development.

Finally, while the work of the Consortium was underway, the world experienced the unprecedented spread of Ebola virus disease across West Africa. During the subsequent global health response it became apparent that the reliance on PCR-based diagnostics resulted in significant delays in diagnosis. Nouvellet *et al.* review the development of rapid diagnostic tests for Ebola virus disease over the past year and use modelling to explore the potential benefits of such tests. Their results demonstrate the key role of rapid diagnostics

to guide individual treatment decisions, while also reducing the potential scale of future epidemics.

Although the outputs of the Diagnostics Modelling Consortium presented in this supplement clearly demonstrate the impact and cost-effectiveness of new diagnostic approaches for multiple diseases of global health significance, the impact of diagnostics remains overlooked. The results can be grave, ranging from overestimation of the impact of interventions when perfect diagnosis is assumed to ignoring the potential role of a diagnostic tool to facilitate lower-cost approaches to treatment. This supplement pulls together a range of articles that highlight the importance of considering the individual- and population-level aspects of the use of diagnostics, encouraging a shift in mindset for all infectious-disease modelling moving forward. In addition, the interdisciplinary nature of this work should not be underestimated. Bringing together the key scientific, clinical and strategic perspectives is imperative from the start of any effort to develop and use technology. Modelling, even at its best, is only a way to describe and quantify our thoughts. To truly determine the impact of diagnostics technologies, they must be evaluated in the field. Only then can we place the appropriate diagnostic and prognostic tools in the right settings to achieve our global health goals.

The editorial process for this supplement was coordinated by Azra Ghani and Alison Reynolds of the Diagnostics Modelling Consortium. We thank Deirdre Hollingsworth, Tim Hallett and Nilufar Hampton for their assistance. We also thank the many anonymous peer reviewers for their support in this process.

1. Strategic Evaluation, Advisory and Development Consulting. *Analysis of POCT/VCT Performed at South African Primary Health Care Clinics* (SEAD, 2011).
2. McNerney, R. & Zumla, A. Impact of the Xpert MTB/RIF diagnostic test for tuberculosis in countries with a high burden of disease. *Curr. Opin. Pulm. Med.* **21,** 304–308 (2015).
3. Rao, V. B., Schellenberg, D. & Ghani, A. C. Overcoming health systems barriers to successful malaria treatment. *Trends Parasitol.* **29,** 164–180 (2013).
4. Hay Burgess, D. C., Wasserman, J. & Dahl, C. A. Global health diagnostics. *Nature* **444** (Suppl 1), 1–2 (2006).
5. United Nations. *The Millenium Development Goals Report* (United Nations, 2015).
6. Smit, P. W. *et al.* Systematic review of the use of dried blood spots for monitoring HIV viral load and for early infant diagnosis. *PLoS ONE* **9,** e86461 (2014).
7. Corran, P. H. *et al.* Dried blood spots as a source of anti-malarial antibodies for epidemiological studies. *Malar. J.* **7,** 195 (2008).
8. Tran, T. M. *et al.* A nested real-time PCR assay for the quantification of *Plasmodium falciparum* DNA extracted from dried blood spots. *Malar. J.* **13,** 393 (2014).
9. Fonjungo, P. N. *et al.* Access and quality of HIV-related point-of-care diagnostic testing in global health programs. *Clin. Infect. Dis.* http://dx.doi.org/10.1093/cid/civ866 (2015).
10. Jani, I. V. *et al.* Effect of point-of-care CD4 cell count tests on retention of patients and rates of antiretroviral therapy initiation in primary health clinics: an observational cohort study. *Lancet* **378,** 1572–1579 (2011).
11. Choko, A. T. *et al.* Uptake, accuracy, safety, and linkage into care over two years of promoting annual self-testing for HIV in Blantyre, Malawi: a community-based prospective study. *PLoS Med.* **12**, e1001873 (2015).
12. World Health Organisation. *Consultation on Technical and Operational Recommendations for Clinical Laboratory Testing, Harmonization and Standardization: Report from the Maputo Meeting* (WHO, 2008).
13. Bill and Melinda Gates Foundation. *Melinda French Gates — Malaria Forum Keynote Address* http://www.gatesfoundation.org/media-center/speeches/2007/10/melinda-french-gates-malaria-forum (Bill and Melinda Gates Foundation, 2007).
14. World Health Organisation. *Global Technical Strategy for Malaria 2016–2030* (WHO, 2015).
15. World Health Organisation. *Accelerating Work to Overcome the Global Impact of Neglected Tropical Diseases: a Roadmap for Implementation* (WHO, 2012).
16. Vlaminck, J., Fischer, P. U. & Weil, G. J. Diagnostic Tools for Onchocerciasis Elimination Programs. *Trends Parasitol.* http://dx.doi.org/10.1016/j.pt.2015.06.007 (2015).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests. Financial support for this publication has been provided by the Bill & Melinda Gates Foundation.

ADDITIONAL INFORMATION

ARTICLE OPEN

Evaluating the impact of pulse oximetry on childhood pneumonia mortality in resource-poor settings

Jessica Floyd*[1], Lindsey Wu*[1,2], Deborah Hay Burgess[3], Rasa Izadnegahdar[3], David Mukanga[3] & Azra C. Ghani[1]

It is estimated that pneumonia is responsible for 15% of childhood deaths worldwide. Recent research has shown that hypoxia and malnutrition are strong predictors of mortality in children hospitalized for pneumonia. It is estimated that 15% of children under 5 who are hospitalized for pneumonia have hypoxaemia and that around 1.5 million children with severe pneumonia require oxygen treatment each year. We developed a deterministic compartmental model that links the care pathway to disease progression to assess the impact of introducing pulse oximetry as a prognostic tool to distinguish severe from non-severe pneumonia in under-5 year olds across 15 countries with the highest burden worldwide. We estimate that, assuming access to supplemental oxygen, pulse oximetry has the potential to avert up to 148,000 deaths if implemented across the 15 countries. By contrast, integrated management of childhood illness alone has a relatively small impact on mortality owing to its low sensitivity. Pulse oximetry can significantly increase the incidence of correctly treated severe cases as well as reduce the incidence of incorrect treatment with antibiotics. We also found that the combination of pulse oximetry with integrated management of childhood illness is highly cost-effective, with median estimates ranging from US$2.97 to $52.92 per disability-adjusted life year averted in the 15 countries analysed. This combination of substantial burden reduction and favourable cost-effectiveness makes pulse oximetry a promising candidate for improving the prognosis for children with pneumonia in resource-poor settings.

Nature **528**, S53–S59 (3 December 2015), DOI: 10.1038/nature16043

Despite interventions being available, it is estimated that pneumonia is responsible for 15% of childhood deaths worldwide[1]. Reductions in annual mortality remain modest, with nearly 950,000 under-5 year olds dying of pneumonia in 2013 (ref. 2). Despite the unprecedented rate of *Haemophilus influenzae* type B (Hib) and pneumococcal vaccine (PCV) introduction, achieving high levels of coverage in developing countries is still challenging[3]. Therefore, in regions where vaccine introduction and scale-up lags behind other countries, improved access to diagnosis and treatment is crucial. This includes interventions at multiple points in the continuum of care — improving care-seeking practices, increasing the availability of suitable diagnostics, and guiding both formal and informal care providers in appropriate disease management. Unfortunately, current treatment coverage remains low, and, more importantly, most childhood pneumonia deaths result from a lack of, or delay in, accurate diagnosis[4].

A crucial component of improving pneumonia outcomes is the early identification of patients at risk of treatment failure and the timely provision of supportive care. However, in the absence of appropriate prognostic tools at the frontline, currently recommended World Health Organization (WHO) guidelines for integrated management of childhood illness (IMCI) often lead to an overuse of antibiotics and the under-referral of patients with severe pneumonia who require hospital care[5]. The most recent 2015 technical update of IMCI guidelines defines non-severe pneumonia as the presence of fast breathing or chest in-drawing or both, which is treatable with oral antibiotics. Severe pneumonia is defined as cough or difficulty breathing in the presence of danger signs, and requires referral to a hospital or health facility for injectable antibiotics or other supportive care such as oxygen therapy[6]. Currently, identification of these IMCI symptoms remains inconsistent and unreliable among community health-care workers or carers without clinical training[7]. Therefore, improved prognostic and diagnostic tools for case-management are necessary to substantially reduce pneumonia-associated morbidity and mortality.

Hypoxaemia and malnutrition are strong predictors of mortality in children who are hospitalized for pneumonia[8,9]. This has led to increasing support for the use of oxygen therapy and monitoring oxygen saturation in the management of severe cases. It is estimated that 15% of children who are hospitalized for pneumonia have hypoxaemia (oxygen saturation, or SpO2, of <90% (ref. 10) and that around 1.5 million children with severe pneumonia require oxygen treatment each year[11]. The use of pulse-oximetry devices (used to measure the oxygen level in the blood) in community health-care settings has been proposed as a method to identify hypoxic children at risk of treatment failure. These devices may be particularly beneficial at the frontline given that they require little training and reduce the reliance on clinical symptoms. The current pulse-oximetry systems are also quick, non-invasive and require minimal infrastructure.

The aim of this study was to evaluate the public health impact and cost-effectiveness of current IMCI guidelines combined with pulse-oximetry devices as a prognostic tool in the hands of frontline health workers in resource-poor settings. To do this, we developed a model of disease progression

*These authors contributed equally. [1]MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, Norfolk Place, London W2 1PG, UK. [2]Department of Immunology and Infection, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. [3]The Bill & Melinda Gates Foundation, 500 Fifth Avenue North, Seattle, Washington 98109, USA. Correspondence should be addressed to: J. F. e-mail: jrf1g15@soton.ac.uk or A. G. e-mail: a.ghani@imperial.ac.uk.

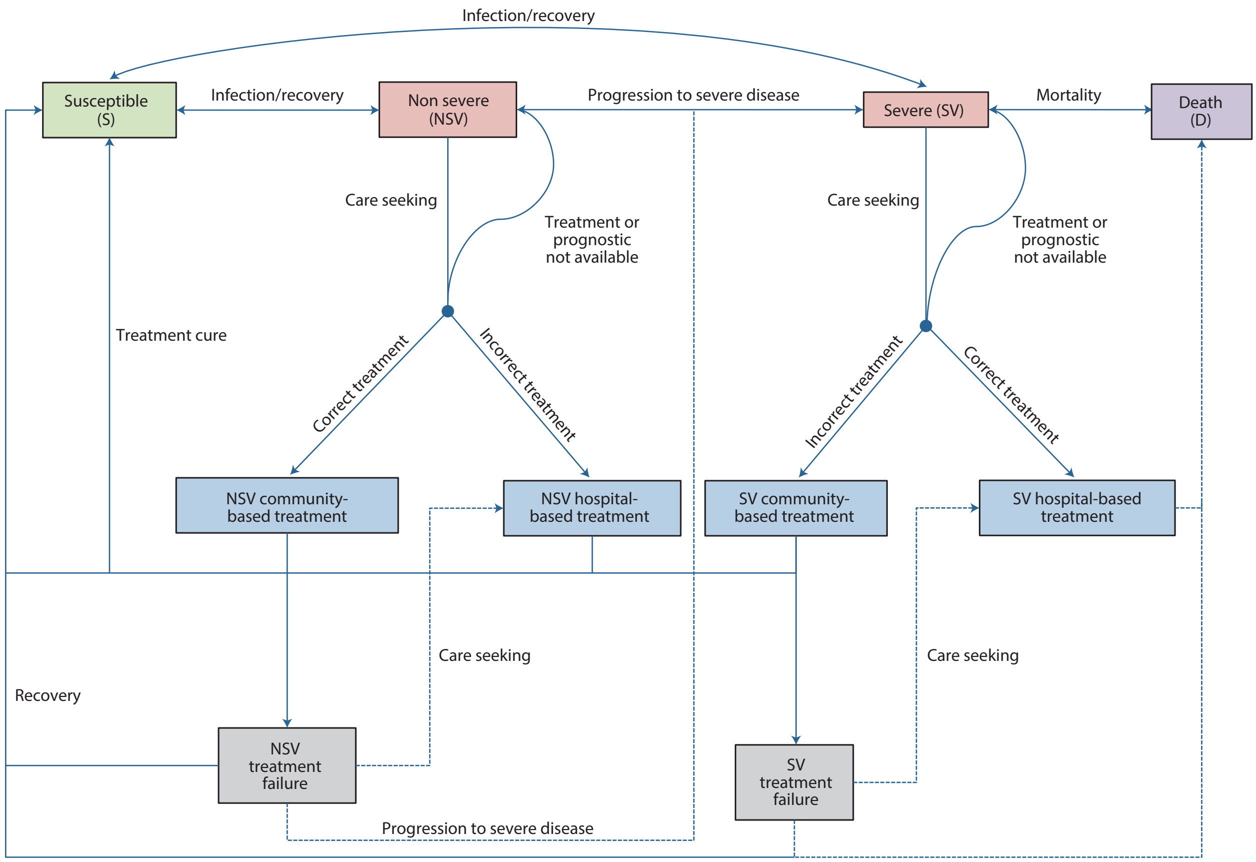


Figure 1 | Overview of model structure. The main states and transitions of the model are shown, with cases transitioning at time-dependent rates. Each state contains a subset of states to track the length of infection.

that explicitly tracks the continuum of care pathways, and parameterized the model for the top 15 countries with the highest burden of pneumonia.

METHODS

Model structure. The progression and treatment of pneumonia in a population of children under the age of five was modelled using a continuous-time deterministic compartmental model (Fig. 1). Full mathematical details of the model are given in the Supplementary Information. Without treatment, children may be in one of four states: susceptible (S), non-severe (NSV), severe (SV) or death (D). These definitions of non-severe and severe disease are different to those defined by IMCI. We classified severe disease as disease that requires hospitalization and non-severe disease as disease that can be successfully managed with oral antibiotics alone. Susceptible children become infected at a constant rate determined by the incidence of pneumonia in the population, and although most enter the non-severe state before progressing, a small proportion progress directly to severe disease. Children with non-severe disease may recover naturally (without treatment), progress to a treatment state or progress to the severe state. Children with severe disease may recover naturally, progress to a treatment state, or die and be tracked in the death state. Each of the non-severe disease and severe disease states are further subdivided into day of infection. Children with non-severe pneumonia who have not progressed to severe disease or recovered naturally after 14 days return to the susceptible state, whereas children with severe pneumonia who have not received treatment or recovered naturally after 14 days (giving a maximum potential illness length of 28 days) are assumed to have died.

As well as making a natural recovery, non-severe and severe cases may separately enter one of two treatment states: treatment received outside of hospital (referred to as community-based treatment) or hospital-based treatment, at rates that depend on the care-seeking rate and factors in the care pathway. The latter is determined by a decision tree in which four factors are included: availability of a prognostic tool, whether or not the prognostic tool gives the correct result, adherence to the prognosis (whether or not the correct treatment, according to the prognosis, was administered by the medical practitioner and followed by the patient) and treatment availability. An example of a section of the decision tree is given in Supplementary Figure 1. We aimed to capture correct and incorrect treatment rates of both non-severe and severe cases, so cases may move into either treatment state depending on the outcome of the decision tree, regardless of whether or not it is the correct treatment. Moreover, this design includes the possibility that cases can move into the correct treatment state even if the prognosis was incorrect.

Last, if treatment fails to work, then the case may move into one of two treatment failure states, determined by a probability of treatment success that depends on both the severity of the disease and whether the treatment is appropriate. For example, severe cases are less likely to be cured by community-based treatment than hospital-based treatment. Children with non-severe disease that fail to respond to treatment may progress to hospital treatment, to a severe state or naturally recover to the susceptible state. Children with severe disease who fail to respond to treatment may progress to hospital treatment, to the death state or naturally recover to the susceptible state at different rates to the non-severe cases.

Community-based treatment is assumed to consist of a course of amoxicillin that lasts for 3 days, whereas hospital-based treatment lasts for 7 days. To ensure that treatment does not prolong illness in the model, those who enter the treatment states at a late stage of infection (such as day 13 or 14 of non-severe illness) are assumed to recover before reaching the end of the treatment, instead of progressing through all 3 or more days of treatment and therefore taking longer to return to susceptible than if they are left untreated (see Supplementary Fig. 2).

Model parameters. The public health impact of the introduction of pulse-oximetry devices was evaluated at the community level in comparison to a baseline standard of care with IMCI by calculating the incremental deaths averted with the introduction of pulse oximetry in different countries. Countries included in

Table 1 | Model parameters. The central estimates shown are derived from the source literature with ranges added for the sensitivity analysis.

Parameter	Value (range)	Sources
Disease progression		
Incidence	Country-specific ± 10%	Ref. 12
Proportion severe on day 1	5% (2–10%)	Ref. 13
Mean duration of non-severe illness before recovery	3 days (2–4 days)	Ref. 25
Mean duration of non-severe illness before progression to severe illness	10 days (9–11 days)	Estimated from model (see Methods)
Mean duration of severe illness before recovery	4 days (3–5 days)	Ref. 22
Mean duration of severe illness before death	7 days (6–8 days)	Ref. 26
Proportion bacterial versus viral (NSV)	85% viral (75–90%) 15% bacterial (25–10%)	Ref. 27
Proportion bacterial versus viral (SV)	85% bacterial (75–90%) 15% viral (25–10%)	Assumed
Care-seeking and health-care parameters		
Mean duration of illness before care seeking	NSV 3 (2–4) days SV 0.75 (0.5–1) days	Ref. 26
Probability that community-based treatment is available	Country-specific ± 10%	Ref. 3
Probability that timely hospital access	0.61 ± 10%	Ref. 24
Probability of community-based treatment curing non-severe bacterial case	0.925 (0.90–0.95)	Ref. 28
Probability of treatment with hospital care curing case	0.925 (0.80–0.95)	Assumed to be high if oxygen is available with lower values representing poorer standard of care
Probability of treatment with amoxicillin curing severe case if prescription adhered to	0.65 (0.6–0.7)	Ref. 29 (based on treatment failure rates of patients with hypoxia at baseline)
Prognostic parameters		
Probability of prognostic available	1 (0.9–1)	Assumed to be high for the purpose of this analysis
Sensitivity of IMCI	0.55 (0.5–0.6)	Ref. 30
Sensitivity of PO1	0.7 (0.65–0.75)	Estimated
Sensitivity of PO2	0.85 (0.8–0.9)	Ref. 14
Specificity of IMCI	0.85 (0.8–0.9)	Assumed to be high given low overall referral rates
Specificity of PO1 and PO2	0.85 (0.8–0.9)	Assumed to be similar to IMCI
Adherence to non-severe prognosis (IMCI)	0.55 (0.5–0.6)	Refs 31–33
Adherence to severe prognosis (IMCI)	0.65 (0.6–0.7)	Refs 31,32
Adherence to non-severe prognosis (PO1 and PO2)	0.55 (0.5–0.6)	Assumed to be similar to IMCI
Adherence to severe prognosis (PO1 and PO2)	0.85 (0.8–0.9)	Assumed to be high for the purpose of this analysis
Prognosed SV treated with community-based treatment versus nothing	1	Assumed that prognosed SV will always be treated even if not referred to hospital
Prognosed NSV that is hospitalized versus receiving nothing	0.025 (0.01–0.05)	Assumed that prognosed NSV are unlikely to be incorrectly hospitalized

IMCI, integrated management of childhood illness; NSV, non-severe pneumonia; PO1, IMCI and pulse oximetry combination with a sensitivity of 70%; PO2, IMCI and pulse oximetry combination with a sensitivity of 85%; SV, severe pneumonia.

Table 2 | Cost parameters. The central estimates are shown derived from the source literature with ranges added for the sensitivity analysis.

Parameter	Value (range)	Sources
Amoxicillin treatment per child	US$0.1614 drug cost plus country-specific delivery cost (± 10%)	Ref. 34
Average hospital cost per episode	Country-specific (± 10%)	Ref. 15
Pulse oximeter	$250	Ref. 35
Batteries	$2 ($1.5–2.5)	Assumed
Uses per set of batteries	840	Ref. 36
Lifetime of device	2 years	Ref. 36
Number of devices needed	1 per 1,000 children under 5 (0.8–1.2)	Assumed
Delivery and distribution costs	20%	Assumed

the analysis were selected according to the number of pneumonia-attributed deaths that occur each year in that country, using estimates from 2011 (ref. 12). Although more recent estimates of pneumonia mortality in under-fives are available, these did not include estimates of the number of cases and so the 2011 estimates were used for the analysis. The 15 countries with the highest number of deaths were chosen, excluding China, Angola and Tanzania owing to a lack of data on the availability of community-based care in these countries.

The incidence of pneumonia in the model was fitted to the mortality data by finding the incidence rate that best matches the mortality data using a normal likelihood. The rate of progression from non-severe to severe disease was fitted to estimates of proportions of cases that progress to severe disease[12,13]. All other disease progression parameters were based on a review of the literature and expert opinion (Table 1). The availability of community-based care was fitted simultaneously to data on the percentage of children with

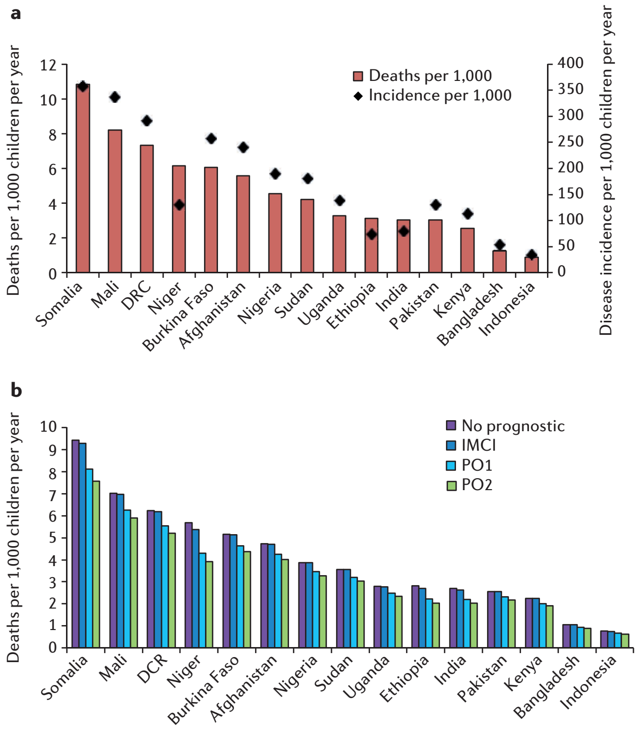


Figure 2 | Pneumonia incidence and mortality. a, Estimated under-5 pneumonia incidence predicted by the model and previously reported mortality[12] in top 15 countries with the highest burden. **b,** Median estimates of pneumonia deaths per 1,000 children under 5 across the 15 countries with the highest mortality for four different community-level prognostics — none, integrated management of childhood illness (IMCI), combined IMCI and pulse oximetry with 70% sensitivity (PO1) and combined IMCI and pulse oximetry with 85% sensitivity (PO2). DRC, Democratic Republic of the Congo.

suspected pneumonia receiving antibiotics[3], assuming a binomial likelihood, whereas the other care-seeking parameters were obtained from the literature where available (Table 1). One care-seeking parameter — the probability of treatment with hospital care curing the case — could not be identified in the literature. For the purpose of assessing the impact of the pulse oximeter as a prognostic tool, we assumed that this was high, representing a situation in which oxygen and other facilities are available. Lower values of this parameter will reduce the impact and cost-effectiveness of any prognostic tool; this was explored in our sensitivity analysis.

To estimate the impact and cost-effectiveness of a new prognostic combination (IMCI and pulse oximetry), we also needed to make a number of assumptions about its availability, its ability to accurately classify a case as severe (sensitivity) or non-severe (specificity) compared with IMCI alone, and adherence to its use. For the purposes of assessing its utility, we assumed it would be made available and hence set this parameter to a high value. Although data were available to support the sensitivity of IMCI, sufficient data were not available to inform the sensitivity of IMCI when combined with pulse oximetry. Thus, we proposed two scenarios: one in which the addition of pulse oximetry increases the sensitivity of IMCI to 70% (referred to as PO1), and one in which the sensitivity of the combination is increased to 85% (referred to as PO2), reflecting the potential of pulse oximetry to identify both people with hypoxic cases and cases with abnormal oxygen saturation (90–95%) who would benefit from referral[14]. We could not find data on specificity and assumed these would be relatively high (85%) for both IMCI alone and the PO1 and PO2 prognostic packages (Table 1). Adherence to IMCI guidelines for both non-severe and severe cases was based on the literature, and we set severe prognosis adherence to the PO1 and PO2 prognostic packages to be higher (increased from 65% to 85%), reflecting the perception that a physical tool would increase the likelihood of adherence. Finally, two parameters are included to account for non-adherence to the prognosis. For those whose prognosis is severe disease, if the mother did not take the child to hospital despite referral we assumed that as a minimum the child would receive amoxicillin (taking into account its availability). For those with a non-severe prognosis, but for whom the treatment regimen was not adhered to, we assumed that they had a high probability of receiving no treatment.

Costing approach. The incremental cost-effectiveness of pulse oximetry was evaluated in comparison to a baseline of using IMCI alone. Costing was undertaken from a public health provider perspective and hence no societal, economic or private sector costs were included.

Costs were subdivided into two categories. Additional direct costs included the cost of prognostic equipment, batteries and delivery and distribution of the prognostic tool. Further health-care costs included additional amoxicillin courses for patients with non-severe disease, increased hospitalization costs for patients with severe disease and the cost savings that would arise from the reduction in inappropriate amoxicillin prescriptions or treatment (owing to assumed increased prognostic adherence) or hospital referral. Total costs are the direct costs plus the additional health-care costs. Wherever possible, costing data were obtained from the literature (Table 2). Health-care costs were obtained at the country level from the WHO–CHOICE database[15]. This included the average cost of an outpatient visit and the cost of inpatient stays based on an average of 7 days of hospitalization. The cost implications arising from potential overuse of amoxicillin (antibiotic resistance) were not included in the analysis given the difficulty of obtaining any meaningful quantitative estimate of this[16]. Disability-adjusted life years (DALYs) are calculated using country-specific life-expectancy values and are not discounted or age-weighted, as suggested by recent recommendations[17]. The overall cost-effectiveness compared with IMCI is then presented as the cost per DALY averted compared with IMCI.

Sensitivity analysis. To explore sensitivity to key parameters, Latin hypercube sampling was used to draw 1,000 different parameter values from a triangular distribution. Outputs from the model were calculated using each set of parameters and the top and bottom 5% of the results were discarded to obtain 90% ranges. Additional sensitivity analyses are reported in the Supplementary Information.

RESULTS

Figure 2a shows previous annual pneumonia mortality estimates[12] and our estimates of the incidence of pneumonia per 1,000 children in the 15 countries with the highest burden. As expected, the estimated pneumonia incidence follows a similar pattern to mortality, but with variations owing to between-country variation in the availability of community-based care. Niger, Ethiopia and India are estimated to have particularly low rates of antibiotic treatment[18], and thus are predicted by the model to have the highest case-fatality rates.

The predicted mortality across the 15 countries under different types of prognostic scenarios are shown in Figure 2b. In comparison to a baseline scenario of no prognostic tool (modelled as a prognostic with a sensitivity and specificity of 50%), IMCI is predicted to have a small incremental impact on mortality. This is largely driven by the fact that the sensitivity of IMCI in identifying severe cases is only 50–60%, and hence not much greater than 'chance'. By contrast, we predict that distribution of pulse oximetry to affected communities could result in much greater reductions in mortality. This is driven by both the higher sensitivity of pulse oximetry (65–75% for PO1, 80–90% for PO2) compared with IMCI (50–60%) and our assumption that there would be a higher adherence rate to PO1 and PO2 (80–90% for severe cases) compared with IMCI (60–70% for severe cases). The countries with the poorest rates of community-based treatment have the greatest reduction in mortality when only IMCI is considered (compared with no prognostic tool); whereas for PO1 and PO2, the countries with the highest incidence of disease are predicted to have the greatest reduction in mortality compared with no prognostic tool used.

The reduction in mortality under IMCI, PO1 and PO2 compared with the absence of a prognostic tool can be translated directly into estimates of deaths averted per 1,000 children at risk. Our estimates of the number of deaths averted by PO1 and PO2 is substantially higher than those estimated to be averted by IMCI, with the highest impact per 1,000 children in Somalia, Mali and Niger (Fig. 3a). Although the effect of IMCI on deaths averted

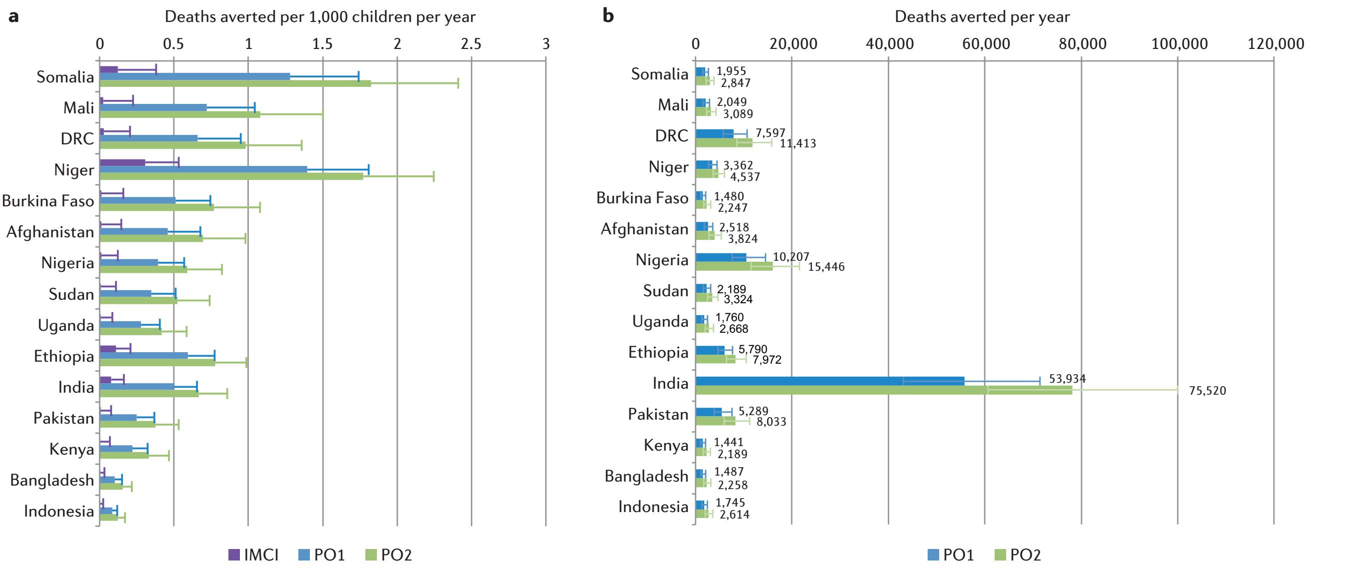


Figure 3 | Deaths averted by prognostic tools across 15 countries. **a**, Estimated deaths averted per 1,000 children per year under integrated management of childhood illness (IMCI), PO1 (combined IMCI and pulse oximetry with 70% sensitivity) and PO2 (combined IMCI and pulse oximetry with 85% sensitivity) each compared with a baseline of no prognostic. **b**, Estimated absolute number of deaths averted by PO1 and PO2 compared with IMCI alone. Error bars show 90% range from the sensitivity analysis. DRC, Democratic Republic of the Congo.

is small relative to pulse oximetry, there is still a benefit compared with the absence of a prognostic tool. The estimated impact of pulse oximetry is more apparent when translated to absolute values scaled to country-specific under-5 population sizes (Fig. 3b). In absolute terms, the introduction of pulse-oximetry devices is estimated to result in the greatest annual reductions in pneumonia deaths in India (75,500 deaths averted per year for PO2) and in Nigeria (15,400 deaths averted per year for PO2) owing to their large under-five populations (128 million and 27 million, respectively). Collectively, we estimate that the implementation of PO1 (the more conservative estimate of IMCI and pulse-oximetry sensitivity) could avert 103,000 (90% range = 77,000–135,000) deaths annually across the 15 countries with the highest burden. For PO2, this increases to 148,000 (90% range = 112,000–193,000).

A key aim of improved prognostic tools is to increase the number of patients with severe disease receiving correct hospital referral. Using Nigeria as a country-level example, we estimated that the proportion of people with severe cases receiving hospital referral could increase by 44% by implementing PO1 or 62% by implementing PO2. We also estimated a substantial reduction in incorrect treatment — with the number of people with severe disease receiving community-based care alone (under treatment) decreasing by 19% (PO1) and 25% (PO2). However, a small increase in number of people with non-severe cases who receive hospital referral (over-treatment) was also predicted (from 3.99% of cases to 5.08% for both PO1 and PO2); this is due to the assumption that higher adherence is associated with a severe prognosis made by PO1 or PO2 compared with IMCI alone.

Increasing the sensitivity of the prognostic tool — which is assumed from adding pulse oximeters to existing IMCI — only has a substantial effect if other aspects of the health system are functioning at reasonable levels. Using Nigeria as a country-level example, we identified four key variables that determined the additional impact of a pulse oximeter — the availability of amoxicillin, hospital care, oxygen within the hospital and the prognostic tool.

The reduction in mortality is slightly more sensitive to the availability of community-based care than the availability of hospital-based care (Fig. 4a). This is due to two assumptions: if a person is referred to hospital, but is unable to access hospital care, then community-based care may instead be accessed; and community-based care includes the provision of a full course of amoxicillin. Therefore, if hospital-based care is unavailable, cases will probably receive amoxicillin instead, and a proportion of these will be cured by amoxicillin. Providing a more sensitive prognostic tool only had a substantial impact on mortality when the availability of hospital care was greater than 20%, when the oxygen availability exceeded 60% and when the prognostic tool was available in more than 60% of communities (Fig. 4a). We further investigated the impact of reduced oxygen availability on both deaths averted and cost-effectiveness (Fig. 4b) and found that oxygen availability (parameterized as the hospital cure rate) needs to be at least 60% for deaths to be averted by PO2 compared with IMCI alone. For cost-effectiveness to be less than US$40 per DALY averted, oxygen availability should exceed 70% (Fig. 4b).

PO1 and PO2 are assumed to have both higher sensitivity and higher adherence to severe prognosis than IMCI alone. To assess the relative contributions of these two parameters, the change in incidence of cases in each of the four treatment states compared with IMCI was calculated for five scenarios: with the increase in adherence to a severe prognosis only; the increase in sensitivity for PO1 only; the PO1 combination of increased adherence and sensitivity; the increase in sensitivity for PO2 only; and the PO2 combination of increased adherence and sensitivity (Supplementary Fig. 3). We found that increasing each parameter alone had a substantial effect on the incidence of deaths and treated cases. Increased adherence to a severe prognosis alone caused more non-severe cases to be incorrectly treated owing to the poor sensitivity of IMCI that caused some non-severe cases to be given a prognosis of severe — these are then treated because of the higher adherence. The increase in severe prognosis adherence resulted in a small increase in incorrect treatment for non-severe cases (relative to the total number of non-severe cases), but a substantial increase in correct treatment for severe cases (relative to the total number of severe cases). The combination of increased adherence and prognostic sensitivity had the largest impact on the correct treatment of severe cases.

Across all the countries studied, community-based care costs (US$0.56–3.70 per course of amoxicillin[15]) are small in comparison to the corresponding hospital costs ($6.44–130.34 for a 7-day inpatient stay[15]). The estimated cost of the intervention itself (approximately $165 per 1,000 children per year) is also estimated to be lower than the additional health-care costs in most of the countries. As such, overall increases in health-care costs under pulse oximetry are largely associated with higher hospital referral rates for severe pneumonia cases.

The cost-effectiveness of implementing pulse oximetry in the 15 countries with the highest burden is shown in Figure 5. PO1 and PO2 are most cost-effective in Niger ($3.72 and $2.97 per DALY, respectively), the Democratic Republic of the Congo ($6.81 and $4.81 per DALY, respectively) and Ethiopia ($6.57 and $5.00 per DALY, respectively), partly driven by the comparatively lower costs of hospital care in those countries. For comparison, an insecticide-treated mosquito net (ITN) to prevent malaria is estimated to cost between $5 and $31 per DALY averted[19], whereas HIV antiretroviral therapy has been estimated to cost upwards of $150 per DALY averted[20].

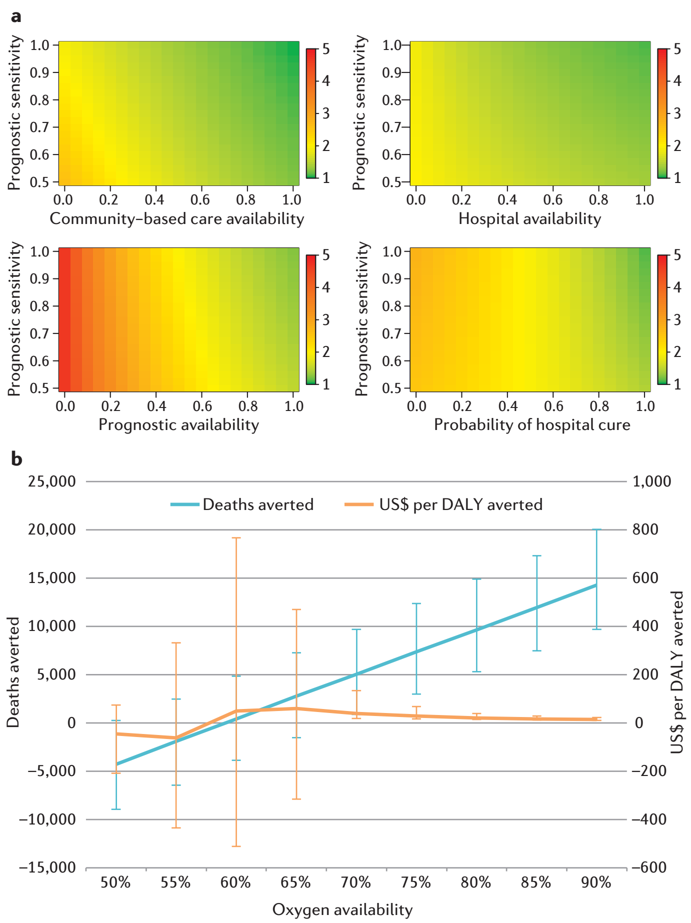


Figure 4 | Sensitivity analyses. a, The relationship between prognostic scenario parameters, prognostic sensitivity and the estimated annual pneumonia deaths per 1,000 children under 5 under varying prognostic scenarios: 1) the availability of amoxicillin (community-based care); 2) the availability of hospital care; 3) the availability of the prognostic tool; and 4) the availability of oxygen (hospital care). All other parameters were fixed at their central values. **b,** The impact of oxygen availability on cost-effectiveness and deaths averted, assuming a combined integrated management of childhood illness and pulse oximetry sensitivity of 85% (PO2). All scenarios are modelled using country-specific parameters for Nigeria. DALY, disability-adjusted life year.

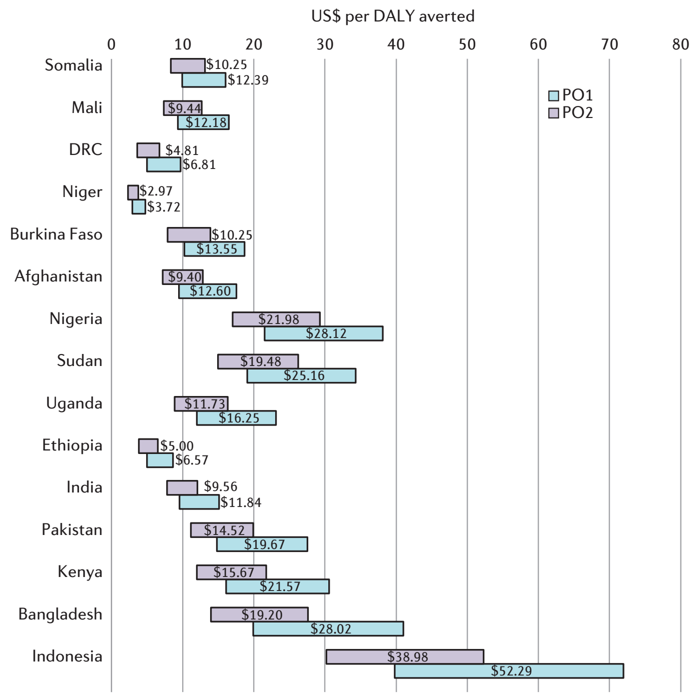


Figure 5 | Cost-effectiveness of prognostic tools. Estimated cost-effectiveness (US$ per disability adjusted life year (DALY)) of PO1 (combined integrated management of childhood illness (IMCI) and pulse oximetry with 70% sensitivity) and PO2 (combined IMCI and pulse oximetry with 85% sensitivity) compared with IMCI in the 15 countries with the highest burden of pneumonia. The numbers indicate the median estimate whereas bars represent the 90% range. DRC, Democratic Republic of the Congo.

DISCUSSION

Using a simple model that links care pathways to the progression of pneumonia in young children, we predict that a combination of pulse oximetry with current IMCI guidelines has the potential to avert up to 148,000 deaths per year in the 15 countries with the highest burden of pneumonia across Africa and Asia, under the assumption that there is more than 90% prognostic tool and supplementary oxygen availability. This equates to about one-sixth of all deaths owing to community-acquired pneumonia in the developing world. For comparison, it has been estimated that complete elimination of low birth weight would prevent 25% of pneumonia deaths in developing countries, with a similar proportion prevented by eliminating malnutrition[21]. Analysis of the impact of the pneumococcal vaccine for infants, PCV10, predicted that the vaccine has the potential to directly avert around 262,000 deaths in under-5s across 72 countries[22]. The relative ease of implementation of a pulse oximetry-based intervention (even with the assumption of perfect availability) compared with the elimination of low birth weight or malnutrition makes it an important candidate for an intervention against pneumonia in resource-poor settings.

On top of the large reduction in deaths, we predict that the addition of pulse oximetry to IMCI has the potential to increase the correct treatment of severe cases by an estimated 44%. When modelling the effect of PO1 and PO2 compared with IMCI, we increased two key parameters to simulate the implementation of pulse oximetry. These were prognostic sensitivity (which was set to be higher for PO1 and PO2 than for IMCI) and adherence to a severe prognostic result (also higher for PO1 and PO2 than for IMCI). Both substantially contribute to an increase in the correct treatment of severe cases and thus the predicted reduction in pneumonia deaths. Sensitivity analysis showed that an increase in either of these characteristics alone has the potential to prevent deaths. These substantial burden reductions are explained by the relatively low sensitivity of IMCI for detecting severe cases (just 55% compared with a potential 70–85% for pulse oximetry combined with IMCI), and the very high burden of pneumonia in these 15 countries (910,000 deaths attributed to pneumonia in under-5s in 2010 (ref. 12).

The incremental cost-effectiveness of PO1 and PO2 over IMCI was found to be very low in 14 of the 15 countries (less than $30 per DALY averted). For reference, the gross domestic product (GDP) per capita across these 14 countries ranged from $400 to $3,000 in 2013. Compared with the cost-effectiveness of the distribution of PCV10, estimated to be $100 per DALY averted[22], this seems to be remarkably favourable. However, the extra costs of providing oxygen support were not taken into account in the calculations of cost-effectiveness, owing to a lack of data on the availability of oxygen support at the country level. Including the costs in the analyses will decrease the cost-effectiveness. Nevertheless, we predict that when coupled with the additional costs of oxygen support, pulse oximetry will still compare favourably with the PCV10 vaccine. For example, a study in Papua New Guinea estimated the cost-effectiveness of improving oxygen support in this area (including oxygen concentrators and the provision of pulse oximeters) to be $50 per DALY averted[23].

There were several limitations to our analysis. One of these was the lack of country-specific data to inform our parameter for access to hospital care. We assumed that 61% of people who were referred to hospital would access hospital care, based on data from a retrospective case review study of children with severe pneumonia in Tanzania[24]. More realistically, we know that the proportion of children reaching an appropriate health facility may vary significantly between countries and so having one single parameter for all countries could result in inaccurate estimates. However, our sensitivity analysis showed

that the model is less sensitive to hospital access than community-based care access. Another limitation was a lack of data on the availability of oxygen support across the 15 countries and how it is distributed throughout the health system. Linked to this was our assumption that the hospital systems in each country were substantial enough to support the extra cases that would be hospitalized. More field data on the hospital systems in each country is required to inform and expand the model to appropriately address these limitations. Nevertheless, it is clear that for any new prognostic to have impact there is the need to also invest in strengthening the existing primary and tertiary healthcare facilities so that appropriate care is provided to those that are referred.

1. World Health Organization/UNICEF. *Ending Preventable Child Deaths from Pneumonia and Diarrhoea by 2025* http://www.who.int/maternal_child_adolescent/documents/global_action_plan_pneumonia_diarrhoea/en/ (WHO, 2015).
2. World Health Organization. *Levels and Trends in Child Mortality 2014* http://www.who.int/maternal_child_adolescent/documents/levels_trends_child_mortality_2014/en/ (WHO, 2015).
3. International Vacine Access Centre. *Pneumonia and Diarrhea Progress Report 2014*. http://www.jhsph.edu/research/centers-and-institutes/ivac/resources/IVAC-2014-Pneumonia-Diarrhea-Progress-Report.pdf (IVAC, 2015).
4. Lim, Y.-W. *et al.* Reducing the global burden of acute lower respiratory infections in children: the contribution of new diagnostics. *Nature* **444,** 9–18 (2006).
5. Izadnegahdar, R., Cohen, A. L., Klugman, K. P. & Qazi, S. A. Childhood pneumonia in developing countries. *Lancet. Respir. Med.* **1,** 574–84 (2013).
6. World Health Organization. *IMCI Chart Booklet* http://www.who.int/maternal_child_adolescent/documents/IMCI_chartbooklet/en/ (WHO, 2015).
7. Sazawal, S. & Black, R. E. Effect of pneumonia case management on mortality in neonates, infants, and preschool children: a meta-analysis of community-based trials. *Lancet Infect. Dis.* **3,** 547–556 (2003).
8. Subhi, R. *et al.* The prevalence of hypoxaemia among ill children in developing countries: a systematic review. *Lancet. Infect. Dis.* **9,** 219–27 (2009).
9. Reed, C. *et al.* Development of the Respiratory Index of Severity in Children (RISC) score among young children with respiratory infections in South Africa. *PLoS ONE* **7,** e27793 (2012).
10. Subhi, R. *et al.* The prevalence of hypoxaemia among ill children in developing countries: a systematic review. *Lancet. Infect. Dis.* **9,** 219–227 (2009).
11. Nair, H. *et al.* Global and regional burden of hospital admissions for severe acute lower respiratory infections in young children in 2010: a systematic analysis. *Lancet* **381,** 1380–1390 (2013).
12. Rudan, I. *et al.* Epidemiology and etiology of childhood pneumonia in 2010: estimates of incidence, severe morbidity, mortality, underlying risk factors and causative pathogens for 192 countries. *J. Glob. Health* **3,** 010401 (2013).
13. Pitt, C., Roberts, B. & Checchi, F. Treating childhood pneumonia in hard-to-reach areas: a model-based comparison of mobile clinics and community-based care. *BMC Health Serv. Res.* **12,** 9 (2012).
14. Madico, G. The role of pulse oximetry. *Arch. Pediatr. Adolesc. Med.* **149,** 1259 (1995).
15. World Health Organization. *Health Service Delivery Costs* http://www.who.int/choice/cost-effectiveness/inputs/health_service/en/ (WHO, 2015).
16. Gandra, S., Barter, D. M. & Laxminarayan, R. Economic burden of antibiotic resistance: how much do we really know? *Clin. Microbiol. Infect.* **20,** 973–980 (2014).
17. Murray, C. J. L. *et al.* GBD 2010: design, definitions, and metrics. *Lancet* **380,** 2063–2066 (2012).
18. UNICEF. *The State of the World's Children 2015: Executive Summary | UNICEF Publications*. http://www.unicef.org/publications/index_77928.html (UNICEF, 2014).
19. Breman, J. G. *et al.* In *Disease Control Priorities in Developing Countries* (eds Jamison, D. T. *et al.* (World Bank, 2006)
20. Alistar, S. S., Grant, P. M. & Bendavid, E. Comparative effectiveness and cost-effectiveness of antiretroviral therapy and pre-exposure prophylaxis for HIV prevention in South Africa. *BMC Med.* **12,** 46 (2014).
21. Victora, C. G. *et al.* Potential interventions for the prevention of childhood pneumonia in developing countries: improving nutrition. *Am. J. Clin. Nutr.* **70,** 309–320 (1999).
22. Sinha, A., Levine, O., Knoll, M. D., Muhib, F. & Lieu, T. A. Cost-effectiveness of pneumococcal conjugate vaccination in the prevention of child mortality: an international economic analysis. *Lancet* **369,** 389–396 (2007).
23. Duke, T. *et al.* Improved oxygen systems for childhood pneumonia: a multihospital effectiveness study in Papua New Guinea. *Lancet* **372,** 1328–1333 (2008).
24. Walter, N. D. *et al.* Why first-level health workers fail to follow guidelines for managing severe disease in children in the Coast Region, the United Republic of Tanzania. *Bull. World Health Organ.* **87,** 99–107 (2009).
25. Hazir, T. *et al.* Comparison of oral amoxicillin with placebo for the treatment of world health organization-defined nonsevere pneumonia in children aged 2–59 months: a multicenter, double-blind, randomized, placebo-controlled trial in pakistan. *Clin. Infect. Dis.* **52,** 293–300 (2011).
26. Källander, K. *et al.* Delayed care seeking for fatal pneumonia in children aged under five years in Uganda: a case-series study. *Bull. World Health Organ.* **86,** 332–338 (2008).
27. Le Roux, D. M., Myer, L., Nicol, M. P. & Zar, H. J. Incidence and severity of childhood pneumonia in the first year of life in a South African birth cohort: the Drakenstein Child Health Study. *Lancet. Glob. Heal.* **3,** e95–e103 (2015).
28. Straus, W. L., Qazi, S. A., Kundi, Z., Nomani, N. K. & Schwartz, B. Antimicrobial resistance and clinical effectiveness of co-trimoxazole versus amoxycillin for pneumonia among children in Pakistan: randomised controlled trial. Pakistan Co-trimoxazole Study Group. *Lancet* **352,** 270–274 (1998).
29. Fu, L. Y. *et al.* Brief hospitalization and pulse oximetry for predicting amoxicillin treatment failure in children with severe pneumonia. *Pediatrics* **118,** e1822–e1830 (2006).
30. Kelly, J. M. *et al.* Community health worker performance in the management of multiple childhood illnesses: Siaya District, Kenya, 1997–2001. *Am. J. Public Health* **91,** 1617–1624 (2001).
31. Chinbuah, M. A. *et al.* Assessment of the adherence of community health workers to dosing and referral guidelines for the management of fever in children under 5 years: a study in Dangme West District, Ghana. *Int. Health* **5,** 148–156 (2013).
32. Acácio, S. *et al.* Under treatment of pneumonia among children under 5 years of age in a malaria-endemic area: population-based surveillance study conducted in Manhica district- rural, Mozambique. *Int. J. Infect. Dis.* **36,** 39–45 (2015).
33. Senn, N. *et al.* Use of antibiotics within the IMCI guidelines in outpatient settings in Papua New Guinean children: an observational and effectiveness study. *PLoS ONE* **9,** e90990 (2014).
34. Management Sciences for Health. *International Drug Price Indicator Guide*. http://erc.msh.org/dmpguide/pdf/DrugPriceGuide_2013_en.pdf (MSF, 2013).
35. Lifebox Foundation. *Lifebox: Saving Lives Through Safer Surgery* http://www.lifebox.org/about-lifebox/our-product/ (Lifebox Foundation, 2015)
36. UNICEF. *Supply Catalogue* https://supply.unicef.org/ (UNICEF, 2015).

SUPPLEMENTARY MATERIAL

Is linked to the online version of this paper at: http://dx.doi.org/10.1038/nature16043

ACKNOWLEDGEMENTS

This study was funded by the Bill & Melinda Gates Foundation Diagnostics Modelling Consortium. L.W. acknowledges doctoral training funding from the UK Medical Research Council (MRC). A.C.G. acknowledges support from the Bill & Melinda Gates Foundation, the UK MRC and the UK Department for International Development.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests. Financial support for this publication has been provided by the Bill & Melinda Gates Foundation.

ADDITIONAL INFORMATION

ARTICLE OPEN

Understanding the incremental value of novel diagnostic tests for tuberculosis

Nimalan Arinaminpathy[1] & David Dowdy[2]

Tuberculosis is a major source of global mortality caused by infection, partly because of a tremendous ongoing burden of undiagnosed disease. Improved diagnostic technology may play an increasingly crucial part in global efforts to end tuberculosis, but the ability of diagnostic tests to curb tuberculosis transmission is dependent on multiple factors, including the time taken by a patient to seek health care, the patient's symptoms, and the patterns of transmission before diagnosis. Novel diagnostic assays for tuberculosis have conventionally been evaluated on the basis of characteristics such as sensitivity and specificity, using assumptions that probably overestimate the impact of diagnostic tests on transmission. We argue for a shift in focus to the evaluation of such tests' incremental value, defining outcomes that reflect each test's purpose (for example, transmissions averted) and comparing systems with the test against those without, in terms of those outcomes. Incremental value can also be measured in units of outcome per incremental unit of resource (for example, money or human capacity). Using a novel, simplified model of tuberculosis transmission that addresses some of the limitations of earlier tuberculosis diagnostic models, we demonstrate that the incremental value of any novel test depends not just on its accuracy, but also on elements such as patient behaviour, tuberculosis natural history and health systems. By integrating these factors into a single unified framework, we advance an approach to the evaluation of new diagnostic tests for tuberculosis that considers the incremental value at the population level and demonstrates how additional data could inform more-effective implementation of tuberculosis diagnostic tests under various conditions.

Nature **528**, S60–S67 (3 December 2015), DOI: 10.1038/nature16045

Every year, nearly three million people develop active tuberculosis (TB), but are not notified to health authorities[1]. Some of these individuals may spontaneously resolve their disease, die or be treated in the private sector, but many remain infectious, fuelling ongoing transmission in the community. Reaching this 'missing three million' remains one of the top priorities for global TB control[2]. A widely cited reason for the ongoing gap between incidence and cases notified is the lack of highly sensitive and deployable diagnostic tests for TB[3]. Sputum smear microscopy, the global cornerstone of TB diagnosis[4], can miss half of all people with infectious TB[5], whereas more sensitive tests cannot routinely be implemented at the point of treatment[6,7]. Nevertheless, the link between improved diagnostic sensitivity and better TB detection remains uncertain. Studies[8–11] in different settings have found little or no change in the number of pulmonary TB diagnoses or deaths when comparing sputum smear microscopy and Xpert MTB/RIF, a more sensitive molecular test[12]. This result may reflect high levels of empirical treatment among people who test negative[13–15]. Against this backdrop, a key question remains: if novel diagnostic tests are developed and implemented at scale, what impact can we expect on TB epidemiology within populations?

The impact of TB diagnostics on transmission reflects not only the accuracy of the test, but also the way in which patients with infectious TB interact with members of the community and with health systems over time[16,17]. These infection pathways have at least three crucial dimensions: the transmission rate (number of transmission events per unit time), the frequency at which people contact health systems (often slower in subpopulations with poor access to care), and the probability of starting effective TB treatment after such contact[18]. Each of these dimensions varies through the duration of infectiousness (from onset to effective treatment, spontaneous recovery or death)[19].

Mathematical models can be a useful tool in helping to demonstrate how these dimensions relate to the impact of diagnostic tests on TB transmission[20–22]. Figure 1a depicts the simplest, and most commonly used[23–25], conceptualization of TB diagnosis in mathematical models so far. In this framework, on becoming infectious, people with TB experience a series of uniform processes. Specifically, they transmit TB at a constant rate, contact the health system at a constant rate and undergo a constant probability of successful diagnosis (leading to appropriate treatment) with each health-system contact. In this framework, the speed at which someone with TB gets treated — and the number of people they infect before that treatment — are strongly related to the sensitivity of the diagnostic algorithm. If, for example, people with TB contact the health system on average every 6 months with a 50% chance of being diagnosed at each visit, the mean duration of infectiousness will be 1 year (approximately the prevalence/incidence ratio estimated by the World Health Organization[1]). If a more sensitive test (for example, replacing sputum smear microscopy with Xpert MTB/RIF[26,27]) can increase that probability of diagnosis from 0.5 to 0.75, the mean duration of disease, and thus the transmission per active case, could be cut by one-third. As a result, the projected epidemiological impact of a more sensitive diagnostic test in this framework is tremendous. This conceptualization of the diagnostic process (constant transmission, constant health-system contact and constant probability of successful diagnosis) over time has permeated nearly all projections of expected epidemiological impact from novel diagnostic tests for pulmonary TB — and it is almost certainly wrong. Figure 1b shows an alternative conceptualization of the TB diagnostic process. In this framework, the transmission rate, frequency of health-system contact and probability of successful diagnosis can all change

[1]MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, Norfolk Place, London W2 1PG, UK. [2]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205, USA. Correspondence should be addressed to: N. A. e-mail: nim.pathy@imperial.ac.uk or D. D. e-mail: ddowdy1@jhmi.edu.

over time[19]. As an illustration, if patients remain infectious for an average of 10 months before seeking care and then begin to contact the health system once a month[28,29], a 50% chance of successful diagnosis per visit would still result in a mean duration of infectiousness of 1 year — but increasing the probability of diagnosis from 0.5 to 0.75 would only reduce that duration to 11.3 months. Worse still, if most transmissions occur in the first 10 months, then even a perfect diagnostic test at the health facility could not avert those events. Thus, the dynamic trajectories of transmission, health-care seeking and diagnostic index of suspicion over the course of TB disease are inextricably linked to the epidemiological impact of novel diagnostic tests[19,30–33] — and overly simple depictions of those trajectories may systematically overestimate that impact. Adding complexity to these simple frameworks requires additional data to inform a more nuanced understanding of the impact of diagnostic tests. Without such data, and models with sufficient flexibility to incorporate them, it is likely that projections of the impact of novel diagnostic tests on TB transmission will continue to be biased, often dramatically so.

So far, test accuracy (sensitivity and specificity) — and to a lesser extent, feasibility of implementation in peripheral settings — has dominated thinking about the 'value' of new TB diagnostic tests. However, the impact of any novel TB diagnostic test will depend on how the health-care system incorporates it[34], as well as on the dynamics of patient interactions with that health-care system (Fig. 1). Epidemiologically, therefore, a novel diagnostic assay should be evaluated not by its sensitivity and specificity, but rather the extent to which it provides diagnostic information beyond earlier tests and practices[35] — its incremental value. This concept is similar to the classic concept of the expected value of diagnostic information (EVDI) promoted by Phelps and Mushlin[36], who also highlighted the need to combine the EVDI with estimates of cost or resource requirements. Subsequent work has expanded on this concept[37,38]. In this paper, we use principles of infectious-disease modelling and diagnostic epidemiology to argue for a change in conceptual approach, from one that has focused primarily on a test's sensitivity to one that centres on its incremental value.

METHODS

Quantifying the incremental value of diagnostic tests for TB. In the context of TB, there are a number of benefits that new diagnostics could provide. These include, but are not limited to, averting TB transmission, averting TB morbidity and mortality[9], saving money[39,40], freeing up health-care capacity for other activities, enabling better treatment of other conditions by ruling out TB[41] and improving patients' economic situations[42] or quality of life[43]. We focus here on the use of novel diagnostic tests as tools to avert TB transmission; however, the intention of some tests may be to add value in one or more of these other areas — and each test's utility should be evaluated according to its intended purpose.

To appropriately estimate the incremental value of a new diagnostic test for TB in terms of transmissions averted, one must consider its relationship to the diagnostic pathways outlined in Figure 1. Table 1 lists four defining features of TB disease and diagnosis (latency[44], gradual symptom onset[45,46], reliance on sputum[47] and concentration of transmission among 'superspreaders'[48]). These features highlight a number of potential diagnostic gaps, or elements along the TB diagnostic pathway, which, if filled by a novel diagnostic test, could generate substantial incremental value.

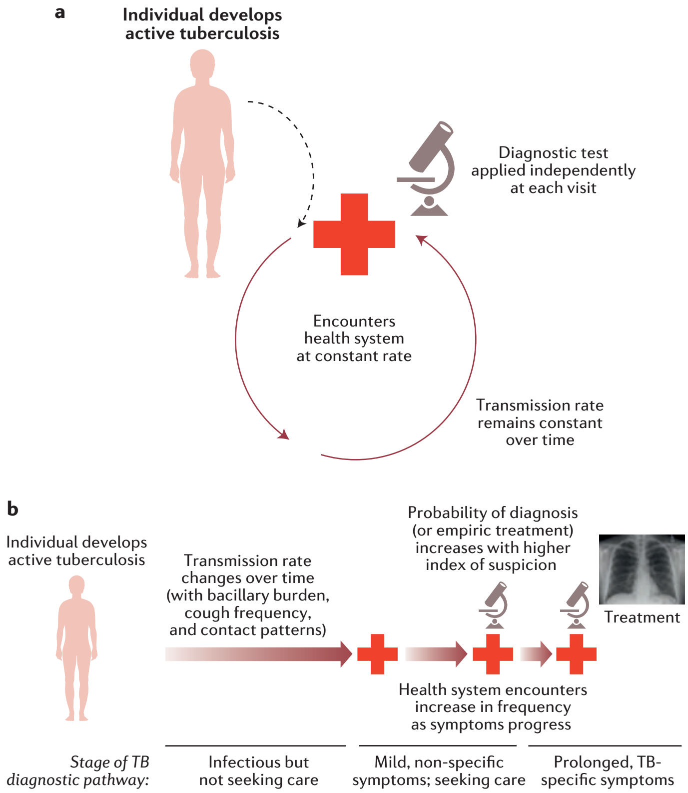


Figure 1 | Conceptual diagrams of different tuberculosis (TB) diagnostic models. **a,** The 'standard' model. So far, most models of TB diagnosis have assumed that, on becoming infectious, individuals with active TB transmit their disease at a constant rate, seek care at a constant rate and maintain a constant probability of diagnosis and treatment with each care-seeking attempt. In reality, the rate at which individuals with active TB transmit disease and seek care, as well as the probability of successful diagnosis and treatment, change over time with the disease course. This process can be more accurately represented by assuming three different stages in the TB diagnostic pathway, as represented at the bottom of **b.** This framework accommodates different types of variation that can be crucial in the potential impact of a test. For example, patients might transmit their disease at an increasing rate over time as bacillary burden increases, seek care more frequently as symptoms progress, and be more likely to receive ancillary diagnostic tests (or empiric treatment) as symptoms persist and other diagnoses become less likely.

Specifically, in any given setting, TB transmission may occur primarily from people who are not sufficiently ill to seek care[49,50]; those who are seeking care, but have symptoms (for example, a mild cough) not specific to TB[51]; or those with severe or prolonged symptoms, but who test negative for TB and are therefore not treated (Fig. 1b). Alternatively, most transmission may occur from hard-to-reach populations in which the rate for seeking care is low[52]. Each of these

Table 1 | Four potential diagnostic gaps in tuberculosis (TB).

Feature of TB natural history	Description	Resultant source of transmission	Potential representation within models	Diagnostic test capable of filling gap
Latency	Prolonged latent period	Individuals who are asymptomatic or have only very mild symptoms	Asymptomatic (or mildly symptomatic) infectious state (I_0)	Test to identify who will progress to active disease, allowing targeted preventive therapy ('progression biomarker')
Slow clinical course	Early non-specific symptoms (for example, cough)	Individuals who are presenting to care, but for syndromic management	Infectious state with symptoms sufficient to drive care seeking, but with low index of suspicion for TB (I_1)	Test to rule out TB (or suggest further testing for TB) in people with a cough ('cough triage test')
Difficult microbiological confirmation	Bacilli often present in low numbers, and only in lungs or sputum; no specific antibody	Individuals who test false negative for TB	Active, care-seeking but undiagnosed state (I_2)	Test to supplant current tests with imperfect sensitivity ('smear replacement test')
Heterogeneous transmission and access to care	Transmission concentrated among those with poor access to care	Individuals who lack sufficient access to seek care rapidly	State with lower care-seeking rate (I')	Smear replacement test for use in peripheral settings with poor access ('point-of-care test')

Table 2 | Profiles of three illustrative diagnostic tests for tuberculosis (TB).

Illustrative name (see Table 1)	Descriptive profile	Mathematical representation	Approximate number needed to screen to identify 1 additional case, typical high-burden setting	Avertable transmission load
Progression biomarker	This test could be applied to a general population to identify people who would subsequently develop active TB; these people could be treated with highly effective preventive regimens	A proportion of individuals with latent TB infection are returned to the uninfected state if successfully identified and treated	100–500 (1/[lifetime probability of incident TB × probability of completing effective preventive therapy])	Pre-care seeking, mild symptoms and prolonged symptoms (general population only)
Cough triage test	This test could be applied to all people presenting to care with a cough, even if suspicion for TB was low — those testing positive could have a highly sensitive test performed	The probability of successful diagnosis in the early symptomatic period increases	20–100 (prevalence of active TB among all patients with a cough)	Mild symptoms and prolonged symptoms (general population only)
Smear replacement test	This test would allow highly sensitive diagnosis among those already seeking care with high suspicion of TB	The probability of successful diagnosis in the late symptomatic period increases	10–20 (prevalence of smear-negative active TB among those with smears currently performed)	Prolonged symptoms (general population only)
Point-of-care test	As for the smear replacement test, but one that is possible to deploy in the poor-access population	As above	10–20 (prevalence of smear-negative active TB among those with smears currently performed)	Prolonged symptoms (low-access population)

gaps suggests a potential diagnostic solution that would have high incremental value. This may be a test to predict progression to active TB (and thus allow targeted preventive therapy), one optimized for diagnosing combinations of symptoms (such as cough and fever), one that is simply more sensitive, or one that is more deployable to peripheral and informal settings (Table 2)[53]. We incorporate these possibilities more formally into a mathematical model of TB transmission.

Model description. Figure 2 presents a simple, illustrative model of TB diagnosis and transmission that expands the constant care-seeking approach shown in Figure 1a. In this model, the population is divided into different compartments that reflect the natural history of TB and incorporate both the stages of the diagnostic pathway shown in Figure 1b and the corresponding diagnostic gaps listed in Table 1. Movement of people between these compartments can be represented by a system of ordinary differential equations, with rates of transition between compartments (for example, γ_0, the rate of initiating care seeking) that reflect the inverse of the mean duration of time spent in each phase (for example, the mean duration between onset of infectiousness and beginning to seek care). As most of these durations are currently unknown (and differ from one setting to the next), we assume — for the purposes of illustration — a population that is at equilibrium, with values of TB incidence, prevalence and mortality that reflect a setting of moderate TB burden (see Supplementary Information). We then use this simplified model to estimate, in this hypothetical setting, the incremental value of diagnostic tests with different profiles under different assumptions about the relative importance of each diagnostic gap. This simplified model divides the population of individuals with active TB into three categories (Figs 1b and 2): those who are infectious, but who are not actively seeking care (I_0), those who have early symptoms that trigger less frequent care seeking and who have a lower probability of correct diagnosis/empiric therapy (I_1), and those who have characteristic and prolonged symptoms that trigger frequent care seeking and a likely diagnosis with each attempt (I_2). We also assume a general population and a sub-population (I', set at 10% for the purposes of illustration) with 'poor' access to care whose rate of care seeking is a specified fraction (k, set initially at 0.5) of the rate in the general population.

Importantly, this model captures the three dynamic processes of transmission, health-care seeking and empiric treatment shown in Figure 1b. First, the rate of transmission (the probability of a 'contact' resulting in TB transmission, multiplied by the number of potential contacts per unit time) can vary over time. For example, β_0 (the number of transmissions per person-month spent in the asymptomatic infectious state I_0) may be higher than β_1 and β_2 (transmission rate from the symptomatic states I_1 and I_2), because the contact rate with susceptible individuals may be highest early in the disease course (suggested by the high prevalence of TB infection in contact investigations[54]). Alternatively, the inverse might be true because the bacillary burden grows over time[55]. We capture this in the concept of the 'transmission load', which we define as the proportion of transmission events at the population level that occur in each of these three stages. Second, the rate of seeking care can increase over time as symptoms progress. Third, the probability of diagnosis with each care-seeking attempt can also increase over time, as symptoms become more suggestive of underlying TB disease[56]. These two processes can be combined into a single 'rate of successful diagnosis and treatment' (d) that increases over time from d_0 to d_1 to d_2.

We explore three hypothetical settings for how transmission varies during the course of TB disease: late diagnostic gap, in which the transmission rate β is four-fold higher at each subsequent stage of TB disease (for example, constant contact rate with susceptible individuals with increasing bacillary burden); early diagnostic gap, in which β falls by a factor of four at each stage (for example, pool of susceptible individuals shrinks over time as household members and other close contacts are exposed); and high access disparity, in which those with least access to care are assumed to have a rate of diagnosis and treatment that is 10% (rather than 50%) that of the general population. Each setting is calibrated to have the same level of TB incidence (see Supplementary Information).

In the context of each of these settings, we explore the potential incremental

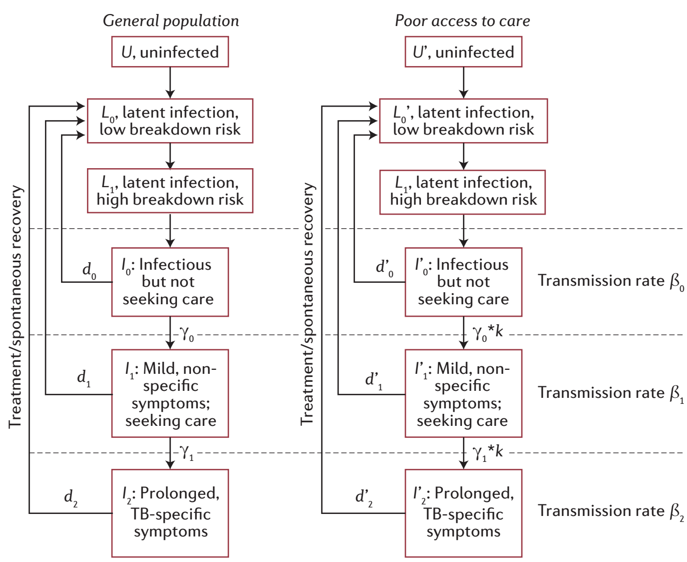


Figure 2 | **Model structure relating diagnostic pathways to transmission load.** A representation of a simple mathematical model that incorporates the three stages of diagnosis shown in Figure 1b. Relative rates of transmission, β, can vary from one stage to the next, with γ representing the inverse of the mean duration of each stage at the population level. Upward arrows denote removal of cases through diagnosis and curative treatment, d, as well as spontaneous resolution (not shown, for simplicity). We also assume a fixed proportion of the population (10% in the base case) have 'poor' access to care, defining an 'access disparity parameter' k to reflect the relative rates of diagnosis in this population. At baseline, we assume that $k = 0.5$. TB, tuberculosis.

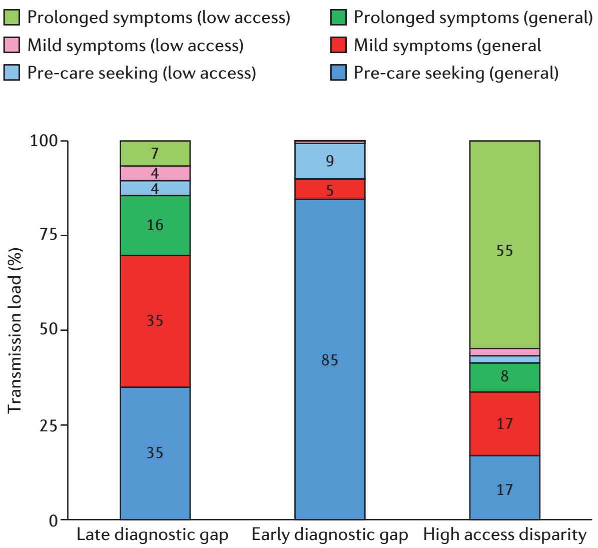


Figure 3 | Tuberculosis (TB) transmission load under three alternative scenarios. The size of each bar denotes the transmission load, defined as the percentage of all tuberculosis transmission that occurs within a given diagnostic stage. Transmission from the general population is shown in darker colours, with that originating from the 'poor-access' population shown in lighter colours. Interrupting transmission at a given stage also averts transmission in subsequent stages (for example, diagnosing a case in stage I_1 also averts the transmission that this case could have caused in stage I_2); this effect can be calculated as the sum of the transmission load in the relevant stage plus all subsequent stages within that population.

value of four illustrative diagnostic tests: a 'progression biomarker' that predicts progression from latent to active TB (to facilitate preventive therapy)[57]; a 'triage' test that facilitates syndromic diagnosis of people presenting with cough[58]; a more sensitive 'replacement test' to supplant current sputum-based confirmatory tests for TB[59]; and a 'point-of-care test' that can replace sputum smear in peripheral settings[60], thereby (unlike the other three tests) being accessible to those with poor access to care. These tests, along with their mathematical representation in our simplified modelling framework, are summarized in Table 2.

We focus on comparisons between these types of diagnostic tests when they are added to the standard of care. To illustrate the transmission contributions of different groups, we assume that progression biomarker, triage and replacement tests are deployed in the general population, whereas the point-of-care test is deployed in the poor-access population. We discuss below how different diagnostic gaps might cause each of these illustrative tests to be preferred over the others, thereby emphasizing the importance of quantifying (or at least estimating) the diagnostic gap in any given setting.

Incorporating resource constraints. Ultimately, discussions of a new diagnostic test's incremental value must also consider any constrained resources — whether economic or otherwise — that would be required to implement the test. One method for evaluating the incremental value of a diagnostic test in a given setting is to first identify any constrained resources required for test implementation. The additional resources required to change from the existing standard of care to an algorithm that augments that standard of care with the new diagnostic test can then be estimated (the incremental resource requirement)[61]. Finally, this is combined with estimates of the incremental number of transmissions averted under this augmented algorithm, relative to the standard of care (incremental impact). Thus, tests that aim to avert TB transmission can be compared using an inverse incremental cost-effectiveness ratio[62]: (incremental transmissions averted)/(incremental resource requirement), or

$$(T_1 - T_0)/(R_1 - R_0) \qquad (1)$$

where 1 denotes the presence of the new test and 0 denotes its absence.

BOX 1 | ESTIMATING THE INCREMENTAL VALUE OF TUBERCULOSIS DIAGNOSTIC TESTS, PER UNIT OF CONSTRAINED RESOURCES.

In comparing diagnostic tests for tuberculosis (TB), it is important to consider both the incremental impact (presented here as transmissions averted) and the incremental resource requirements associated with implementing each new test. The following considerations are not meant to be an exhaustive list, but a demonstration of some of the complexity that must be considered (and corresponding data collected) to properly evaluate the incremental value of diagnostic tests for TB in the setting of constrained resources. Illustrative considerations therefore include:

Determinants of incremental impact (incremental transmissions averted)

1. Epidemiological setting/existing diagnostic gaps
2. Diagnostic test characteristics (accuracy, diagnostic gap targeted)
3. Existing diagnostic algorithms (incremental role of the new test)

Determinants of incremental resource requirements

1. Enumeration of constrained resources
2. Number of tests needed to identify one additional case
3. Per-test outlay of constrained resources ('unit cost')

In settings in which TB diagnostic tests are being compared with other interventions (for example, TB treatment or HIV diagnosis), transmissions averted can be converted into measures of health utility (such as disability-adjusted life years, or DALYs, averted)[63] to estimate resources in terms of economic costs and to report this incremental value as an incremental cost-effectiveness ratio. However, when only comparing diagnostic tests with the same primary aim (to avert transmission), the formulation of incremental value in Equation 1 may be more useful; this formulation places the emphasis on impact rather than cost and does not require additional model assumptions to convert transmissions into DALYs or constrained resources (for example, human capacity) into economic costs. Therefore, we use this more direct formulation in our model results.

RESULTS

Incremental value of TB diagnostic tests. Figure 3 shows how the transmission load at equilibrium (the proportion of population-level transmission contributed by each stage) differs in each transmission scenario. For example, in the late diagnostic gap scenario, 35% of all transmission originates from individuals with mild symptoms in the general population, whereas this percentage falls to 5% in the early diagnostic gap scenario. Importantly, averting transmission in the earlier stages (for example, preventing a case from developing, even before to care seeking) also averts that transmission in later stages — seen in Figure 3 by the combined value of the stacked bars. Thus, for example, preventing all transmission in the latter two care-seeking stages in the general population would avert 51% (35% + 16%) of all transmission in the late diagnostic gap scenario, compared with only 5% in the early diagnostic gap scenario — and a diagnostic test targeting these stages (for example, the 'cough triage' test) might be expected to have greater impact in settings that more closely resemble the late diagnostic gap scenario.

A notable feature of the late diagnostic gap scenario is that, despite transmission being substantially more intense[64] in the prolonged-symptom stage I_2(16 times greater per unit time than in the pre-care-seeking stage I_0), the contribution of this stage to transmission remains relatively modest. This is largely due to the relatively short time that individuals spend in this late symptomatic stage. We assume here that, under the standard of care (typically using sputum smear microscopy), individuals are diagnosed on average after 1 month in this late symptomatic stage, compared with 6 months spent in the asymptomatic stage. However, the high access disparity scenario shows the

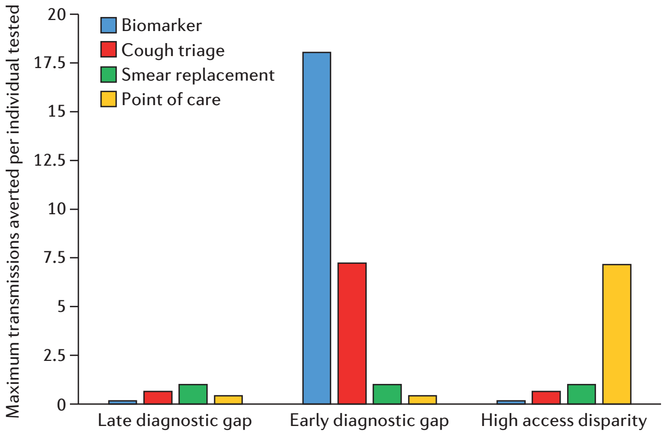


Figure 4 | Maximum incremental impact per unit of constrained resources for four illustrative diagnostic tests in three alternative scenarios. On the *y*-axis is the maximum incremental impact (number of tuberculosis (TB) transmissions averted) for each of four illustrative TB diagnostic tests, divided by the incremental resources required to implement each test (Equation 1). All measures are benchmarked to an incremental impact of 1 for the smear replacement test in the general population. Here, we assume that the constrained resources are simply proportional to the number of people needed to test to diagnose one additional case of active TB. The maximum incremental impact is the number of transmissions that would be averted if diagnosis averts all transmission associated with a given patient stage in Figure 2. Accordingly, the results presented here should be interpreted as an upper bound that are illustrative of the role of diagnostic gaps in each stage. In the cases illustrated here, the 'progression biomarker' (which identifies individuals at risk for progression to active TB) is clearly favoured in the early diagnostic gap scenario, whereas the point-of-care test (which replaces the smear test in the poor-access population, and is deployed only in the poor-access population) is strongly favoured in the high access disparity scenario.

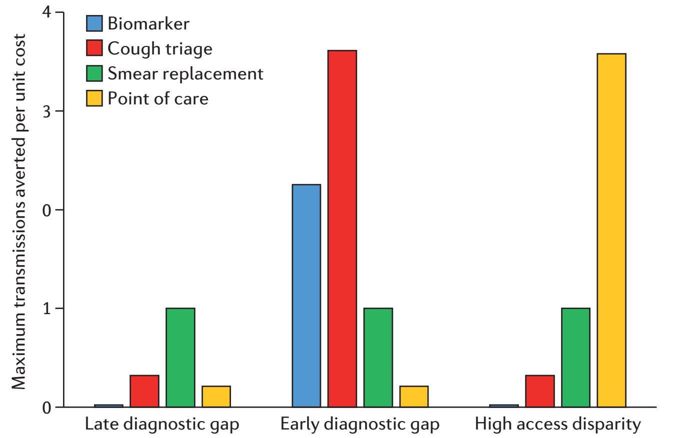


Figure 5 | Maximum incremental value per unit of constrained resources after incorporation of a cost function. The same illustrative tests are evaluated in the same alternative scenarios as in Figure 4, but in this case we apply a cost function that accounts for the fact that diagnosis earlier in the disease process, or among low-access populations, is generally more resource-intensive on a per-test basis (see Supplementary Table 2 for full assumptions). After considering this cost function, the 'progression biomarker' is no longer clearly favoured in the early diagnostic gap scenario, and the degree to which the point-of-care test is favoured over the smear replacement test in the high access disparity scenario is reduced by the same factor (8 in this case) by which the cost per person screened in the low-access population exceeds that in the general population. As in Figure 4, all measures are benchmarked to an incremental impact of 1 for the smear replacement test in the general population.

potential importance of the late symptomatic stage when the rate of diagnosis is diminished. Here, transmission in the late symptomatic stage is sufficiently strong for 55% of the transmission load to occur from a high-risk (and symptomatic) subgroup that accounts for no more than 10% of the total population — a level of disproportionate transmission that is only modestly higher than has been suggested in some settings[65].

Incremental value of new diagnostic tests under constrained resources. Figure 4 shows results for the incremental value (Equation 1), comparing diagnostic tests that target different stages and under different transmission scenarios. For the denominator of Equation 1, Figure 4 assumes a simple, illustrative example for which the constrained resource is the number of individuals who can be tested with a novel test, irrespective of the test type or its unit cost (see Supplementary Table 2 for further details). This might, for example, reflect a setting in which donor funding could be obtained to implement a new test, but the equipment or human resources available to conduct those tests were extremely limited. For the numerator of Equation 1, Figure 4 assumes the maximum number of transmissions averted if the diagnostic test in question (such as the cough triage test) could avert all of the transmission occurring in the stage of disease targeted (for example, I_1, mild symptoms, but seeking care). In practice, owing to factors such as imperfect sensitivity and incomplete population-level implementation, an actual test would only avert a portion of that maximum transmission load; the actual incremental value of each test would therefore be proportionally lower. Thus, in dividing the maximum incremental impact by the fixed incremental resources available, Figure 4 compares the maximum incremental value for each idealized test type, leaving it to subsequent work to estimate what proportion of that maximum could actually be achieved by a given test in practice.

Figure 4 illustrates that — where the primary diagnostic gap is early in the disease course — the maximum incremental value for tests that target earlier stages is higher than that of the smear-replacement test. By contrast, when the primary diagnostic gap is late in the disease course, the maximum incremental value of the later-stage diagnostics is far greater (as represented by their markedly higher incremental value). Notably, where transmission is concentrated among a population with particularly poor access to care, the maximum incremental value for a test that can be implemented in this population can be considerably higher than for any other test (as in the access disparity scenario).

Figure 5 shows an alternative scenario for the denominator of Equation 1 in which the limiting resource is financial (for example, a fixed amount of money available), assuming that the cost per test is higher when applied earlier in the diagnostic pathway. (For example, it is more costly to screen a patient for TB in a prevalence survey[66] than it is in a clinic[67].) The 'unit cost' of a test is also assumed to be higher per person when applied in the poor-access population (see Supplementary Table 2), as these individuals are assumed to be harder to reach than the general population. In the early diagnostic gap scenario, for example, considering this unit cost dramatically lowers the maximum incremental value of the biomarker test that could be achieved per unit of the constrained resource, relative to the cough triage and smear replacement tests. As a result, under this alternative resource constraint, the cough triage test, rather than the biomarker, would be preferred.

DISCUSSION

In evaluating novel diagnostic tests for TB, it is crucial that we move beyond simple considerations of elements such as sensitivity, specificity and turnaround time — and instead begin to consider the incremental value of diagnostic tests that fit certain profiles. We use a simple mathematical model to demonstrate key trade-offs in an illustrative setting. This work demonstrates how diagnostic tests for TB can be quantitatively assessed in terms of their incremental value (incremental impact divided by incremental resource requirement), and moreover how this incremental value can vary from one setting to the next.

The prevailing diagnostic gap in a given setting has a profound effect on the potential incremental impact of each diagnostic test. When most transmission occurs before patients begin to seek care, diagnostic tests that require patients to access the health system are unlikely to have substantial epidemiological impact; thus, in the early diagnostic gap scenario (Fig. 3) the only diagnostic test capable of averting the bulk of the transmission load is the prevention biomarker. Similarly, when a substantial disparity exists between

the high-risk and general population, diagnostics that cannot be implemented in the high-risk group are limited in their potential.

Consideration of incremental impact must also include consideration of incremental resource requirements, however. For example, the resources required to avert a transmission are generally much greater when diagnostic tests are performed early in the disease course[66], or in hard-to-reach populations. As a result, those diagnostic tests with the largest maximum incremental impact may also be those that require the most resources. In estimating the incremental resource requirement of a given test, it is important to consider the resources for TB control that are constrained in a given setting. In many cases, these constrained resources will be purely financial, but in others, there may be limitations on the availability of trained staff or laboratory capacity to perform certain tests[68]. The per-test incremental outlay of the most constrained resources is therefore also likely to vary from one setting to the next. Ultimately, the incremental value of a TB diagnostic test depends not on sensitivity and specificity, but also on multiple factors that will vary from one system to the next (Box 1). For any setting, all six of the elements in Box 1 should be evaluated to help to identify the type of novel test that is likely to have the greatest incremental value (avert the most TB transmission events, given the constrained resources). As assessments of these factors are performed across a variety of settings, consensus may emerge as which tests should be prioritized for development.

Unfortunately, we currently lack the empirical data in most settings to make such an informed assessment. Specifically, it is likely that different transmission loads and diagnostic gaps — early, late or among high-risk subpopulations — predominate in different settings, and that resource constraints vary widely from one setting to the next. How can this data gap be closed?

First, we require better evidence regarding how novel diagnostic tests function when implemented under field conditions. Such data would allow us to estimate the proportion of any diagnostic gap that a new TB test could close, as well as the number of tests required to make one additional diagnosis. Unfortunately, most diagnostic tests are evaluated primarily in well-funded trials and demonstration studies, without good evidence of how they perform in the real world. For example, Xpert MTB/RIF was recommended on the basis of high-quality data about its accuracy and cost-effectiveness under controlled conditions and in a large field trial[26]; however, emerging evidence has suggested that, in many settings, the characteristics of Xpert may be different when implemented in the field — including its sensitivity[69], calibration[70], positive predictive value (owing to low pre-test probability)[71], and accuracy for rifampin resistance[72]. To make accurate assessments of the incremental value of diagnostics, we should collect such data early after launch, and update expectations and recommendations as those data become available.

Second, we need better data on the performance of existing tests, including clinical judgement. These data would enable us to evaluate the incremental number of transmissions that a novel test might be able to avert, relative to the existing standard of care. A series of recent high-quality studies suggests that, when patients present with symptoms that are highly suggestive of TB in upper-middle income settings (for example, South Africa and Brazil), the probability of empirical diagnosis is reasonably high[8–11] — but that a large number of people may be presenting to care with a cough without TB ever being considered[8]. Such studies are crucial to understand the likely diagnostic gaps for TB, but unfortunately, very few such analyses have been performed in settings with fewer resources (for example, most of sub-Saharan Africa[73,74] and Southeast Asia) where empirical diagnosis rates (and the capacity to implement novel diagnostic tests) may be much lower. Characterizations of relative TB transmission from high-risk populations (akin to the 'low-access' population in Figure 3) compared with the general population are also sparse[75], and could potentially be informed by better use of surveillance data[76].

Third, and perhaps most challengingly, we need to prioritize characterizations of the transmission load and diagnostic gaps in a variety of settings. If we can describe the prevailing transmission loads in any given setting, we can then quantify the maximum incremental impact (transmissions averted) of any diagnostic test in that setting. Ultimately, for any setting, one should be able to delineate what proportion of the transmission load in each of the phases of TB (pre-care seeking, mildly symptomatic and prolonged symptomatic in the general population and in high-risk groups) is being averted using existing tests, and therefore what proportion might still be amenable to implementation of a novel diagnostic. Molecular characterization of TB (for example, through whole-genome sequencing[77]) in entire populations is becoming available and can be linked to conventional epidemiological investigations (for example, through contact investigations[78]) using increasingly discriminatory tools for analysis and data collection[79]. Thus, it may become possible to triangulate an infectious individual's onset of symptoms, initiation of care-seeking activities and specific transmission events. Studies that merge data on transmission, contact patterns, symptom histories, care-seeking patterns and interactions with the health-care system on a population level should be prioritized in this regard. In the meantime, simple investigation of surveillance data can help to identify geographic hotspots of transmission, and operational analyses of diagnostic test implementation can demonstrate where diagnoses are probably being missed. Although estimating the duration of an infectious episode poses significant challenges, household cohorts using currently available tools could cast some light on the 'transmission load' that occurs early in the clinical course[80,81].

Finally, we need better investigations of constrained resources in specific settings to enumerate the resources that are genuinely constrained, and to quantify those resources per test performed (as the equivalent of a unit cost). Although conventional economic evaluations of interventions against diseases such as TB implicitly consider money to be the most constrained resource, other studies in low-income settings have shown that human resources, laboratory capacity, regulatory infrastructure or ability to implement new interventions may be the key limiting factors[68]. This may be especially true in the modern era of direct assistance for health — which may supply money, but not resources in the form of trained personnel[82]. An understanding of the most constrained resources in any given setting must then be merged with data on the number of tests required to identify an incremental case, as well as the per-test resource outlay, for any given novel diagnostic test. Only if we truly understand the resources that are most constrained in a given setting, as well as the resource outlay for each type of diagnostic test, can we identify the diagnostic tests that will optimize epidemiological impact under existing resource constraints.

Ultimately, the only way to end TB is to diagnose and treat people with TB before transmission occurs — novel diagnostics are an essential component of any strategy with this aim. If we are to succeed in that endeavour, we must think of, and quantify, those tests not just in terms of sensitivity, specificity and turnaround time, but rather in terms of their incremental value across a variety of epidemiological settings. We present a framework for estimating this incremental value that also highlights the need for additional data in order to inform more appropriate prioritization of novel TB diagnostic tests, across settings that may differ in their existing diagnostic gaps and resource constraints. As we continue to develop diagnostic tests with the goal of curbing TB transmission, we must think beyond accuracy and consider the broader context of patient behaviour, health systems and TB natural history.

1. World Health Organization. *Global Tuberculosis Report, 2014* (WHO, 2014).
2. Herbert, N. *et al.* World TB Day 2014: finding the missing 3 million. *Lancet* **383,** 1016–1018 (2014).
3. Lawn, S. D. *et al.* Advances in tuberculosis diagnostics: the Xpert MTB/RIF assay and future prospects for a point-of-care test. *Lancet Infect. Dis.* **13,** 349–361 (2013).
4. TB CARE I. *International Standards for Tuberculosis Care* 3rd Edn (TB CARE I, 2014).
5. Steingart, K. R. *et al.* Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *Lancet Infect. Dis.* **6,** 570–581 (2006).
6. Cobelens, F. *et al.* Which new diagnostics for tuberculosis, and when? *J. Infect. Dis.* **205,** S191–S198 (2012).
7. Pho, M. T. *et al.* Optimizing tuberculosis case detection through a novel diagnostic device placement model: the case of Uganda. *PLoS ONE* **10,** e0122574 (2015).
8. Churchyard, G. *et al.* Effect of Xpert MTB/RIF on early mortality in adults with suspected TB: a pragmatic randomized trial. *Proc. Conference on Retroviruses and Opportunistic Infections* (IAS-USA, 2014).
9. Theron, G. *et al.* Feasibility, accuracy, and clinical effect of point-of-care Xpert MTB/RIF testing for tuberculosis in primary-care settings in Africa: a multicentre, randomised, controlled trial. *Lancet* **383,** 424–435 (2014).
10. Durovni, B. *et al.* Impact of replacing smear microscopy with Xpert MTB/RIF for diagnosing tuberculosis in Brazil: a stepped-wedge cluster-randomized trial. *PLoS Med.* **11,** e1001766 (2014).
11. Sachdeva, K. S. *et al.* Use of Xpert MTB/RIF in Decentralized Public Health Settings and Its Effect on Pulmonary TB and DR-TB Case Finding in India. *PLoS ONE* **10,** e0126065 (2015).

12. Steingart, K. R. *et al*. Xpert(R) MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database Syst. Rev.* **1,** CD009593 (2013).
13. Theron, G. *et al*. Do high rates of empirical treatment undermine the potential effect of new diagnostic tests for tuberculosis in high-burden settings? *Lancet Infect. Dis.* **14**, 527–532 (2014).
14. Sun, A. Y., Denkinger, C. M. & Dowdy, D. W. The impact of novel tests for tuberculosis depends on the diagnostic cascade. *Eur. Respir. J.* **44,** 1366–1369 (2014).
15. Menzies, N. A., Cohen, T., Murray, M. & Salomon, J. A. Effect of empirical treatment on outcomes of clinical trials of diagnostic assays for tuberculosis. *Lancet Infect. Dis.* **15,** 16–17 (2015).
16. Dye, C. The potential impact of new diagnostic tests on tuberculosis epidemics. *Indian J. Med. Res.* **135,** 737–744 (2012).
17. Langley, I. *et al*. Assessment of the patient, health system, and population effects of Xpert MTB/RIF and alternative diagnostics for tuberculosis in Tanzania: an integrated modelling approach. *Lancet Glob. Health.* **2,** e581–e591 (2014).
18. Dowdy, D. W. & Chaisson, R. E. The persistence of tuberculosis in the age of DOTS: reassessing the effect of case detection. *Bull. World Health Organ.* **87,** 296–304 (2009).
19. Dowdy, D. W., Basu, S. & Andrews, J. R. Is passive diagnosis enough? The impact of subclinical disease on diagnostic strategies for tuberculosis. *Am. J. Respir. Crit. Care Med.* **187,** 543–551 (2013).
20. Zwerling, A. *et al*. Modeling of novel diagnostic strategies for active tuberculosis - a systematic review: current practices and recommendations. *PLoS ONE* **9,** e110558 (2014).
21. Garnett, G. P., Cousens, S., Hallett, T. B., Steketee, R. & Walker, N. Mathematical models in the evaluation of health programmes. *Lancet* **378,** 515–525 (2011).
22. Dowdy, D. W., Dye, C. & Cohen, T. Data needs for evidence-based decisions: a tuberculosis modeler's 'wish list'. *Int. J. Tuberc. Lung Dis.* **17,** 866–877 (2013).
23. Dye, C., Garnett, G. P., Sleeman, K. & Williams, B. G. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. *Lancet* **352,** 1886–1891 (1998).
24. Dowdy, D. W., Chaisson, R. E., Moulton, L. H. & Dorman, S. E. The potential impact of enhanced diagnostic techniques for tuberculosis driven by HIV: a mathematical model. *AIDS* **20,** 751–762 (2006).
25. Abu-Raddad, L. J. *et al*. Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics. *Proc. Natl Acad. Sci. USA.* **106,** 13980–13985 (2009).
26. Boehme, C. C. *et al*. Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study. *Lancet* **377,** 1495–1505 (2011).
27. Boehme, C. C. *et al*. Rapid molecular detection of tuberculosis and rifampin resistance. *N. Engl. J. Med.* **363,** 1005–1015 (2010).
28. Storla, D. G., Yimer, S. & Bjune, G. A. A systematic review of delay in the diagnosis and treatment of tuberculosis. *BMC Public Health* **8,** 15 (2008).
29. Sreeramareddy, C. T., Panduru, K. V., Menten, J. & Van den Ende, J. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature. *BMC Infect. Dis.* **9,** 91 (2009).
30. Salje, H. *et al*. The importance of implementation strategy in scaling up Xpert MTB/RIF for diagnosis of tuberculosis in the Indian health-care system: a transmission model. *PLoS Med.* **11,** e1001674 (2014).
31. Sachdeva, K. S. *et al*. The potential impact of up-front drug sensitivity testing on India's epidemic of multi-drug resistant tuberculosis. *PLoS ONE* **10**, e0131438 (2015).
32. Lin, H. H., Dowdy, D., Dye, C., Murray, M. & Cohen, T. The impact of new tuberculosis diagnostics on transmission: why context matters. *Bull. World Health Organ.* **90,** 739–747A (2012).
33. Mandal, S. & Arinaminpathy, N. Transmission modeling and health systems: the case of TB in India. *Int. Health* **7,** 114–120 (2015).
34. McNerney, R., Cunningham, J., Hepple, P. & Zumla, A. New tuberculosis diagnostics and rollout. *Int. J. Infect. Dis.* **32,** 81–86 (2015).
35. Grobbee, D. E. & Hoes, A. W. *Clinical Epidemiology: Principles, Methods, and Applications for Clinical Research* (Jones and Bartlett, 2009).
36. Phelps, C. E. & Mushlin, A. I. Focusing technology assessment using medical decision theory. *Med. Decis. Making* **8,** 279–289 (1988).
37. Laking, G., Lord, J. & Fischer, A. The economics of diagnosis. *Health Econ.* **15,** 1109–1120 (2006).
38. Van den Bruel, A., Cleemput, I., Aertgeerts, B., Ramaekers, D. & Buntinx, F. The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. *J. Clin. Epidemiol.* **60,** 1116–1122 (2007).
39. van't Hoog, A. H. *et al*. Optimal triage test characteristics to improve the cost-effectiveness of the Xpert MTB/RIF assay for TB diagnosis: a decision analysis. *PLoS ONE* **8,** e82786 (2013).
40. Millman, A. J. *et al*. Rapid molecular testing for TB to guide respiratory isolation in the U.S.: a cost-benefit analysis. *PLoS ONE* **8,** e79669 (2013).
41. Cobelens, F., van Kampen, S., Ochodo, E., Atun, R. & Lienhardt, C. Research on implementation of interventions in tuberculosis control in low- and middle-income countries: a systematic review. *PLoS Med.* **9,** e1001358 (2012).
42. Tanimura, T., Jaramillo, E., Weil, D., Raviglione, M. & Lonnroth, K. Financial burden for tuberculosis patients in low- and middle-income countries: a systematic review. *Eur. Respir. J.* **43,** 1763–1775 (2014).
43. Bauer, M., Leavens, A. & Schwartzman, K. A systematic review and meta-analysis of the impact of tuberculosis on health-related quality of life. *Qual. Life Res.* **22,** 2213–2235 (2013).
44. Getahun, H., Matteelli, A., Chaisson, R. E. & Raviglione, M. Latent *Mycobacterium tuberculosis* infection. *N. Engl. J. Med.* **372,** 2127–2135 (2015).
45. Corbett, E. L. *et al*. Epidemiology of tuberculosis in a high HIV prevalence population provided with enhanced diagnosis of symptomatic disease. *PLoS Med.* **4,** e22 (2007).
46. Mao, T. E. *et al*. Cross-sectional studies of tuberculosis prevalence in Cambodia between 2002 and 2011. *Bull. World Health Organ.* **92,** 573–581 (2014).
47. Ho, J., Marks, G. B. & Fox, G. J. The impact of sputum quality on tuberculosis diagnosis: a systematic review. *Int. J. Tuberc. Lung Dis.* **19,** 537–544 (2015).
48. Gardy, J. L. *et al*. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med.* **364,** 730–739 (2011).
49. Ayles, H. *et al*. Prevalence of tuberculosis, HIV and respiratory symptoms in two Zambian communities: implications for tuberculosis control in the era of HIV. *PLoS ONE* **4,** e5602 (2009).
50. Hoa, N. B. *et al*. Yield of interview screening and chest X-ray abnormalities in a tuberculosis prevalence survey. *Int. J. Tuberc. Lung Dis.* **16,** 762–767 (2012).
51. Banda, H. T. *et al*. Prevalence of tuberculosis in TB suspects with short duration of cough. *Trans. R. Soc. Trop. Med. Hyg.* **92,** 161–163 (1998).
52. Lonnroth, K. *et al*. Towards tuberculosis elimination: an action framework for low-incidence countries. *Eur. Respir. J.* **45,** 928–952 (2015).
53. Denkinger, C. M. *et al*. Defining the needs for next generation assays for tuberculosis. *J. Infect. Dis.* **211** (Suppl 2), S29–S38 (2015).
54. Morrison, J., Pai, M. & Hopewell, P. C. Tuberculosis and latent tuberculosis infection in close contacts of people with pulmonary tuberculosis in low-income and middle-income countries: a systematic review and meta-analysis. *Lancet Infect. Dis.* **8,** 359–368 (2008).
55. Lawn, S. D., Kerkhoff, A. D. & Wood, R. Progression of subclinical culture-positive tuberculosis to symptomatic disease in HIV-infected individuals. *AIDS* **25,** 2190–2191 (2011).
56. Harries, A. D. *et al*. Management of pulmonary tuberculosis suspects with negative sputum smears and normal or minimally abnormal chest radiographs in resource-poor settings. *Int. J. Tuberc. Lung Dis.* **2,** 999–1004 (1998).
57. Wallis, R. S. *et al*. Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice. *Lancet* **375**, 1920–1937 (2010).
58. Garcia-Basteiro, A. L. & Cobelens, F. Triage tests: a new priority for tuberculosis diagnostics. *Lancet Respir. Med.* **3,** 177–178 (2015).
59. Kik, S. V., Denkinger, C. M., Jefferson, C., Ginnard, J. & Pai, M. Potential market for novel tuberculosis diagnostics: worth the investment? *J. Infect. Dis.* **211** (Suppl 2), S58–S66 (2015).
60. Pai, N. P. & Pai, M. Point-of-care diagnostics for HIV and tuberculosis: landscape, pipeline, and unmet needs. *Discov. Med.* **13,** 35–45 (2012).
61. Adang, E., Voordijk, L., Jan van der Wilt, G. & Ament, A. Cost-effectiveness analysis in relation to budgetary constraints and reallocative restrictions. *Health Policy* **74,** 146–156 (2005).
62. Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J. & Stoddart, G. L. In *Methods for the Economic Evaluation of Health Care Programmes* 396 (Oxford Univ. Press, 2005).
63. Azman, A. S., Golub, J. E. & Dowdy, D. W. How much is tuberculosis screening worth? Estimating the value of active case finding for tuberculosis in South Africa, China, and India. *BMC Med.* **12,** 216 (2014).
64. Behr, M. A. *et al*. Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. *Lancet* **353,** 444–449 (1999).
65. Dowdy, D. W., Golub, J. E., Chaisson, R. E. & Saraceni, V. Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics. *Proc. Natl Acad. Sci. USA.* **109,** 9557–9562 (2012).
66. Glaziou, P. *et al*. Tuberculosis prevalence surveys: rationale and cost. *Int. J. Tuberc. Lung Dis.* **12,** 1003–1008 (2008).
67. Schnippel, K. *et al*. Scaling up Xpert MTB/RIF technology: the costs of laboratory- vs. clinic-based roll-out in South Africa. *Trop. Med. Int. Health* **17,** 1142–1151 (2012).
68. Travis, P. *et al*. Overcoming health-systems constraints to achieve the Millennium Development Goals. *Lancet* **364,** 900–906 (2004).
69. Sohn, H. *et al*. Xpert MTB/RIF testing in a low tuberculosis incidence, high-resource setting: limitations in accuracy and clinical impact. *Clin. Infect. Dis.* **58,** 970–976 (2014).
70. Cowan, J. *et al*. Implementing rapid testing for tuberculosis in Mozambique. *Bull. World Health Organ.* **93,** 125–130 (2015).
71. Page-Shipp, L. *et al*. Successes, challenges and lessons from a novel deployment of Xpert((R)) MTB/RIF at a major South African public event. *Int. J. Tuberc. Lung Dis.* **18,** 438–440 (2014).
72. Kelly, J. D. *et al*. Xpert MTB/RIF false detection of rifampin-resistant tuberculosis from prior infection. *Am. J. Respir. Crit. Care Med.* **190,** 1316–1318 (2014).
73. Nakiyingi, L. *et al*. Clinical predictors and accuracy of empiric tuberculosis treatment among sputum smear-negative HIV-infected adult TB suspects in Uganda. *PLoS ONE* **8,** e74023 (2013).
74. Yoon, C. *et al*. Impact of Xpert MTB/RIF testing on tuberculosis management and outcomes in hospitalized patients in Uganda. *PLoS ONE* **7,** e48599 (2012).
75. Dowdy, D. W., Azman, A. S., Kendall, E. A. & Mathema, B. Transforming the fight against tuberculosis: targeting catalysts of transmission. *Clin. Infect. Dis.* **59,** 1123–1129 (2014).
76. Jenkins, H. E. *et al*. Assessing spatial heterogeneity of multidrug-resistant tuberculosis in a high-burden country. *Eur. Respir. J.* **42,** 1291–1301 (2013).
77. van Soolingen, D. Whole-genome sequencing of *Mycobacterium tuberculosis* as an epidemiological marker. *Lancet Respir. Med.* **2,** 251–252 (2014).
78. Walker, T. M., Monk, P., Smith, E. G. & Peto, T. E. Contact investigations for outbreaks of *Mycobacterium tuberculosis*: advances through whole genome sequencing. *Clin. Microbiol. Infect.* **19,** 796–802 (2013).
79. Merker, M. *et al*. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nature Genet.* **47,** 242–249 (2015).
80. Grandjean, L. *et al*. Transmission of multidrug-resistant and drug-susceptible tuberculosis within households: a prospective cohort study. *PLoS Med.* **12,** e1001843 (2015).

81. Zelner, J. L. *et al*. Age-specific risks of tuberculosis infection from household and community exposures and opportunities for interventions in a high-burden setting. *Am. J. Epidemiol.* **180,** 853–861 (2014).
82. Ravishankar, N. *et al*. Financing of global health: tracking development assistance for health from 1990 to 2007. *Lancet* **373,** 2113–2124 (2009).

SUPPLEMENTARY MATERIAL

Is linked to the online version of this paper at: http://dx.doi.org/10.1038/nature16045

ACKNOWLEDGEMENTS

The work of D.D. was partially funded by the B. Frank and Kathleen Polk Assistant Professorship in Epidemiology at the John Hopkins Bloomberg School of Public Health.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests. Financial support for this publication has been provided by the Bill & Melinda Gates Foundation.

ADDITIONAL INFORMATION

ARTICLE OPEN

Sustainable HIV treatment in Africa through viral-load-informed differentiated care

Working Group on Modelling of Antiretroviral Therapy Monitoring Strategies in Sub-Saharan Africa*

There are inefficiencies in current approaches to monitoring patients on antiretroviral therapy in sub-Saharan Africa. Patients typically attend clinics every 1 to 3 months for clinical assessment. The clinic costs are comparable with the costs of the drugs themselves and CD4 counts are measured every 6 months, but patients are rarely switched to second-line therapies. To ensure sustainability of treatment programmes, a transition to more cost-effective delivery of antiretroviral therapy is needed. In contrast to the CD4 count, measurement of the level of HIV RNA in plasma (the viral load) provides a direct measure of the current treatment effect. Viral-load-informed differentiated care is a means of tailoring care so that those with suppressed viral load visit the clinic less frequently and attention is focussed on those with unsuppressed viral load to promote adherence and timely switching to a second-line regimen. The most feasible approach to measuring viral load in many countries is to collect dried blood spot samples for testing in regional laboratories; however, there have been concerns over the sensitivity and specificity of this approach to define treatment failure and the delay in returning results to the clinic. We use modelling to synthesize evidence and evaluate the cost-effectiveness of viral-load-informed differentiated care, accounting for limitations of dried blood sample testing. We find that viral-load-informed differentiated care using dried blood sample testing is cost-effective and is a recommended strategy for patient monitoring, although further empirical evidence as the approach is rolled out would be of value. We also explore the potential benefits of point-of-care viral load tests that may become available in the future.

Nature 528, S68–S76 (3 December 2015), DOI: 10.1038/nature16046

Monitoring people on antiretroviral therapy (ART) cost-effectively is crucial for the sustainability of ART programmes in sub-Saharan Africa. In most countries, patients are required to attend clinics every 1 to 3 months for clinical assessment. The cost of which — for personnel, infrastructure and maintenance — is comparable with costs of the antiretroviral drugs themselves[1–3]. In most settings, patients are monitored by a CD4 count measurement every 6 months with clinical observation at least every 3 months, but they are rarely switched to second-line regimens. A reduction in visit frequency for patients who do not require an adherence intervention or a switch to second-line ART would benefit programmes by reducing costs and benefit patients by saving travel costs and time away from work, possibly lowering the rate of default from care[4].

The biomarker that most directly measures the ongoing effect of ART is the HIV RNA level in plasma (the viral load). If viral load is suppressed, this indicates that the patient is adhering to the drug regimen and does not carry drug-resistant virus. Data from high-income countries suggest that after 1–2 years of ART with viral-load suppression the visit frequency can be reduced. If the viral load is not suppressed this suggests that there is a need for improved adherence and/or a switch in regimen. In most countries in sub-Saharan Africa, measurement of viral load is not widely available. Quantification of HIV RNA requires sophisticated facilities and skilled staff and the costs have been high, although costs have substantially decreased in the past 5 years[5,6]. Modelling studies have indicated that there is a benefit to viral-load monitoring compared with monitoring strategies based on the CD4 count or clinical observation[7–16], but viral-load monitoring has not been found to be cost-effective[7,10–14], owing to the cost of viral-load tests and second-line regimens. Currently, the most feasible approach to begin to measure viral load in many countries is to collect samples as dried blood spots (DBS). DBS are stable at ambient temperature and can be prepared from capillary whole blood, eliminating the need for phlebotomy services[15]. Using existing networks for early infant HIV diagnosis, they can be transported to a regional or national laboratory with results subsequently returned to the clinic by, for example, mobile phone text messaging. However, the presence of cells and low sample volume in DBS specimens means that sensitivity and specificity for detecting whether the level is above the 1,000 copies per millilitre threshold that is used to define viral suppression are imperfect and it is unclear if the approach is adequate[5,16–27]. Looking to the future, it is anticipated that point-of-care (POC) tests — tests that enable a decision to be made about patient management during the same visit that the sample is taken — may become widely available[28], and this may result in greater accuracy than the use of DBS, as well as facilitating rapid action based on the test result.

In the light of these issues, we consider how HIV treatment programmes in low-income countries in sub-Saharan Africa should monitor patients on ART in a way that is likely to lead to the greatest population health gains from the limited resources available[29]. We update a model previously used to compare monitoring strategies, incorporating new lower costs and the potential for viral-load-informed ‘differentiated care’ based on reducing clinic visit costs by reducing visit frequency among virally-suppressed individuals[30,31].

METHODS

The HIV Synthesis transmission model is an individual-based stochastic model of heterosexual transmission, natural history, clinical disease and treatment of HIV infection that incorporates use of specific drugs, resistance mutations and adherence[8,32–36].

*List of working group members and their affiliations appear at the end of the paper. Correspondence should be addressed to: A. P. e-mail: andrew.phillips@ucl.ac.uk.

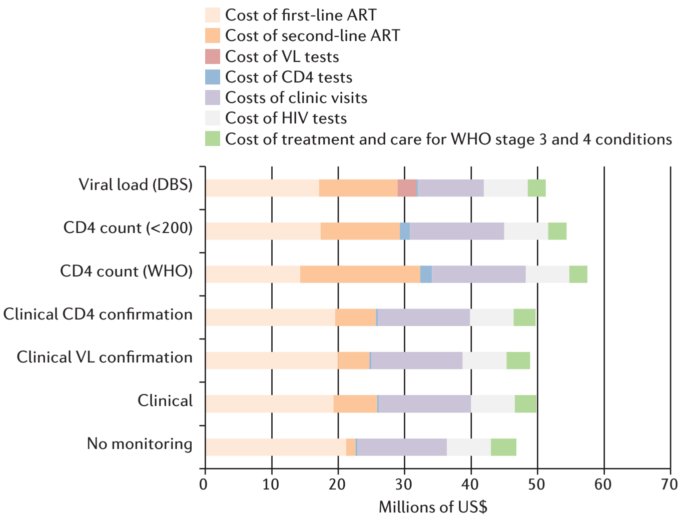


Figure 1 | Overall programme costs. Costs in US$m per 3 months, according to monitoring strategy (mean 2015–2034, discounted at 3% per annum from 2015). ART, antiretroviral therapy; VL, viral load; WHO, World Health Organization.

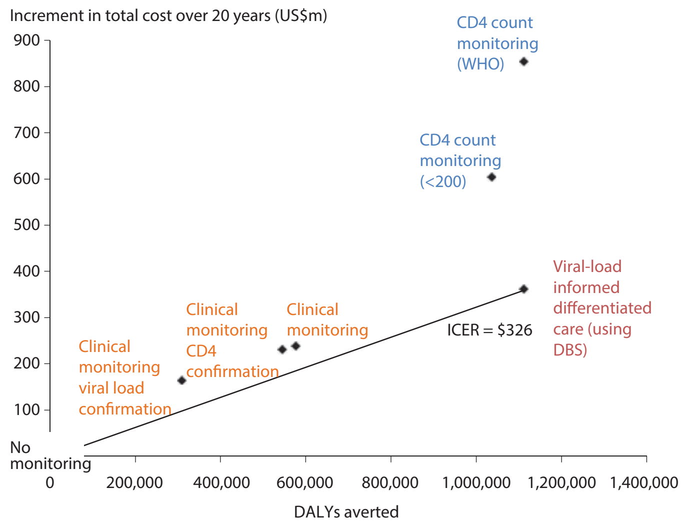


Figure 2 | Cost-effectiveness. Cost-effectiveness plane showing clinical- and CD4-based monitoring strategies along with viral-load-informed differentiated care using dried blood spots. DALYs, disability-adjusted life years; ICER, incremental cost-effectiveness ratio.

ART programme and monitoring strategy modelling

We based our simulated population around that of Zimbabwe and the underlying model is described in detail in the Supplementary Information. We assumed that up to 2015 a CD4 count monitoring strategy has been used. Then we considered the introduction of plausible alternative monitoring strategies and predicted outcomes over 20 years to 2035. The seven main monitoring strategies compared (Table 1) cluster into three main types: clinical observation (with or without targeted CD4 count or viral-load testing in those with clinical disease), regular CD4 count monitoring or regular viral-load monitoring. In the case of viral-load monitoring, we simulate a strategy consisting of off-site laboratory-based testing of DBS using the World Health Organization (WHO) recommended 1,000 RNA copies (cps) ml^{-1} threshold. Viral load measured as <1,000 cps ml^{-1} in the past year is assumed to lead to a reduction in non-ART programme costs owing to fewer clinic visits by people on first-line ART. Measurement of viral load 1,000 cps ml^{-1} or more is assumed to lead to a targeted adherence intervention, which increases adherence in some people. We refer to this strategy as viral-load-informed differentiated care. Regardless of the monitoring strategy used, once strategy-specific failure criteria are met we assume a probability of switching to a second-line regimen of 0.5 per 3 months. In practice, current switch rates are lower than this, even in settings with viral-load monitoring in place[37–39]; we chose this higher probability, however, to be able to discern differences in effects between strategies. In sensitivity analyses we consider a situation in which switch rates are zero. Throughout, we assume monitoring is performed only for people on first-line ART.

We model decreased precision of DBS for measuring viral load by considering the presence of HIV RNA in cells and the small sample volume[5,25,40], such that the sensitivity and specificity of the measure for detecting viral load of >1,000 cps ml^{-1} compared with measurement on a plasma sample are 86% and 92%, respectively (compared with values ranging from 81% to 85% sensitivity and 88% to 99% specificity[5] for most assays); we consider other values in sensitivity analysis. We also assume that there is a 3-month delay in the clinician acting on the result, even though results are generally returned to the clinic quicker than this.

Sensitivity analyses were performed to consider: possible differences in population adherence profile, potential increases in sexual behaviour, changes in effectiveness of the adherence intervention triggered by viral load being >1,000 cps ml^{-1}, a policy of initiation of ART at diagnosis, that visit frequency might be reduced in those with a CD4 count of >350 per mm^3 in the past year, a zero rate of switch to second-line regimens, differences in the baseline prevalence of HIV, differences in the proportion on ART, differences in the rate of ART interruption if visit frequency has been reduced owing to viral load being <1,000 cps ml^{-1}, a higher discount rate of 5% rather than 3%, and a 10-year rather than a 20-year time horizon. In addition, we considered the effects of whether whole blood or plasma is used, whether the test is done in a central laboratory and incurs the 3-month delay in action or is done at POC with no delay, the threshold to define failure (200, 1,000 or 5,000 cps ml^{-1}, which is only assessed in the context of plasma), and the frequency of measurement (every 6 months, annually or every 2 years).

Last, we focussed on the specific comparison between viral load using DBS and using a plasma-based POC test to quantify the extent of various potential advantageous features of a POC test on its cost-effectiveness in relation to use of DBS. It is important to note that we are considering potential features of a POC test — it is not clear that such features can be delivered, so this analysis is directed mainly towards developers and should not be interpreted as indicating that POC tests will prove to have any of these advantageous features. This is why we chose to consider a plasma-based POC test, although in reality it may be more likely that a whole-blood-based test is used for many POC tests, to avoid a plasma separation step. Further details of how all these aspects are modelled are provided in the Supplementary Information.

Economic analysis

Our objective is to maximize population health — the health benefits associated with the alternative monitoring strategies are estimated using the metric disability-adjusted life years (DALYs) averted — with the available resources. A health sector perspective has therefore been adopted for the analysis. Direct and indirect costs incurred by the patients are excluded. Both costs and health benefits were discounted to present value using a 3% per annum discount rate in our base case. The expected costs and health outcomes associated with each monitoring strategy can be compared to indicate which is likely to represent the best value from the available resources. The cost-effectiveness threshold for a country represents the opportunity costs of resources required to fund the intervention, in terms of the health gains that those resources could generate if used for alternative purposes in the public health-care system[41]. As such, the threshold for a country is not readily apparent, but US$500 per DALY averted is likely to be at the upper end based on the magnitude of benefit if the resources were spent on other programmatic priorities such as eliminating coverage gaps for ART if these are large[42]. The modelling results are intended to inform decisions in sub-Saharan African countries such as Zimbabwe classified as low and low-middle income using the World Bank country classifications that have typically struggled to scale-up viral-load monitoring[31]. The analyses may also be informative for higher income countries in the region (such as South Africa and Botswana) that have already scaled up viral-load monitoring, but are seeking more efficient ways to deliver ART.

Table 1 | The seven main monitoring strategies modelled listed by the short name given to the strategy.

	No monitoring	Clinical monitoring	Clinical monitoring viral load confirmation	Clinical monitoring CD4 count confirmation	CD4 count monitoring (WHO)	CD4 count monitoring (<200)	Viral-load-informed differentiated care using DBS
What the monitoring strategy entails for people on first-line ART	NA	Check on presence of symptoms every 3 months	Check on presence of symptoms every 3 months Measure viral load if WHO 4 condition diagnosed or two WHO 3 conditions diagnosed in 1 year	Check on presence of symptoms every 3 months Measure CD4 count if WHO 4 condition diagnosed or two WHO 3 conditions diagnosed in 1 year.	6-month CD4 count If failure criteria seem to be met, re-measure to confirm (confirmatory CD4 count)	12-month CD4 count If failure criteria seem to be met, re-measure to confirm (confirmatory CD4 count)	Viral load measure using DBS at 6 months, 12 months and every 12 months thereafter If viral load is >1,000 cps ml^{-1} then provide adherence intervention and re-measure viral load 3 months later (confirmatory viral load measure) No CD4 count measurements
Failure criteria	NA	WHO 4 condition diagnosed or two WHO 3 conditions diagnosed in 1 year	Viral load >1,000 cps ml^{-1}	CD4 count <250 mm^{-3}	CD4 count less than pre-ART baseline or CD4 count <100 mm^{-3} in confirmatory CD4 count	CD4 count <200 mm^{-3} after more than 3 years on ART. CD4 <100 mm^{-3} after more than 1 year on ART in confirmatory CD4 count	Viral load >1,000 cps ml^{-1} in confirmatory viral load measure
Reduction in clinical visit frequency and hence reduction in non-ART programme cost*	None	None	None	None	None	None	Yes, when most recent viral load <1,000 cps ml^{-1}, measured in the past year

*Assuming 3-monthly clinical visits for all strategies except under viral-load-informed differentiated care when the most recent viral load <1,000 cps ml^{-1}, measured in past year. More frequent clinical visits than once every 3 months are not modelled as the model advances in 3-month periods. ART, antiretroviral therapy; Cps, copies; DBS, dried blood spot; WHO 4, World Health Organization stage 4 condition.

Disability weights to calculate DALYs averted were derived from a recent comprehensive study[43]. Unit costs (in US$ at 2014 prices) are detailed in the Supplementary Information. In brief, costs of viral-load assays are assumed to be $22. This is a fully-loaded cost, counting all components such as reagents, costs of equipment, human resources, buildings, and so on (see Supplementary Information). Because POC viral-load tests are not yet available it was not possible to calculate the cost so we assumed a similar cost of $22, although it is likely that the fully-loaded cost will be higher than this. The cost of measuring CD4 counts is assumed to be $10 (ref. 44). The current annual cost (including supply chain) of the first-line regimen of efavirenz, emtricitabine and tenofovir (assumed to be used as a fixed-dose combination) is assumed to be $144 per person per year and for the second-line regimen of zidovudine, emtricitabine and ritonavir-boosted atazanavir to be $312 per person per year[45]. Annual programme costs for clinic visits (not including drug, or viral load or CD4 count tests) are $80 per year[1,2], with an assumed reduction to $40 per year, after measurement of viral suppression because of reduced clinical visit frequency of every 6 months from every 1 to 3 months (with interim pharmacy-only visits, depending on the amount of drug that can be dispensed).

RESULTS

The status of the simulated population in 2014 is shown in Table S1 in Modelling Methods in the Supplementary Information. Mean predicted outcomes over 20 years are shown in Table 2. The proportion of people who are taking or have taken ART (ART-experienced), who have fulfilled the criteria for failure of first-line ART is lowest with no monitoring and is below 15% for each of the clinical monitoring strategies. It is highest for the CD4 count monitoring (WHO) strategy (41%) because the failure definition is fulfilled if the CD4 count is below the pre-ART baseline level (which can occur due to high CD4 count variability, and particularly if ART has been interrupted). The proportion is intermediate for the CD4 count monitoring (<200) strategy and viral-load-informed differentiated care using DBS strategies (at 26% and 27%, respectively). The proportion of all people on ART who have viral suppression is highest with the viral-load-informed differentiated care using DBS strategy (86%) and lowest with no monitoring (76%), with the small range of 10% reflecting the generally high levels of adherence (although we consider in sensitivity analyses a situation in which adherence levels are lower and the proportion with viral suppression is accordingly lower). The death rate is markedly lower for the CD4 count and viral-load monitoring strategies than for the other strategies, and this is particularly evident in those among whom viral-load failure has occurred. Notably, there is also a benefit of viral-load-informed differentiated care using DBS on HIV incidence over all the other strategies.

Costs and their components by monitoring strategy are shown in Figure 1. Programme costs for clinic visits are lowest with viral-load-informed differentiated care using DBS owing to the reduction in clinic visit frequency among virally-suppressed people. Figure 2 shows the cost-effectiveness plane, showing the total incremental DALYs averted in the population over 20 years, together with the incremental costs (both discounted), compared with no monitoring. Owing to the higher death rate of people on ART and higher HIV incidence, the clinical monitoring strategies avert fewer DALYs than the viral load and CD4-count-based monitoring strategies. Additional costs incurred are highest for CD4-count monitoring, particularly the CD4 count monitoring (WHO) strategy. Viral-load-informed differentiated care using DBS averts a similar number of DALYs as CD4-count monitoring and is the most cost-effective strategy owing to the reduction in non-ART programme costs in people with viral suppression, with an incremental cost-effectiveness ratio (ICER) of $326 per DALY averted. Figure 3 depicts how the cost-effectiveness is affected by the assumed costs of viral-load tests and savings in clinic visit costs in people with suppressed viral load. In our base case viral-load test cost of $22, viral-load-informed differentiated care is cost-effective only if reduced clinic visits provide at least a $30 per person per year saving offset.

The effect of varying model assumptions are shown in Figure 4 and Supplementary Figure 1. Changes in the sensitivity and specificity of viral-load measurement using whole blood (as used for DBS) did not markedly influence the ICER, nor did the extent of the assumed effect of viral-load measurement >1,000 cps ml^{-1} on adherence. The ICER for viral-load-informed differentiated care was lower when we assumed lower population adherence and when we assumed higher population levels of unprotected sex, resulting in higher HIV incidence. In a scenario with a switch rate of zero, viral-load-informed differentiated care was cost saving. Confirming the results shown in Figure 3, if no reduction in visit frequency is assumed with viral-load monitoring (Supplementary Fig. 1u) then it is not cost-effective. The only other scenarios in which viral-load-informed differentiated care was not cost-effective was when we considered a 10-year time horizon instead of 20 years and when we considered a doubling of rate of ART interruption in people with a reduced visit frequency owing to viral load being <1,000 cps ml^{-1} (Figure 4 and Supplementary Fig. 1q and r).

Table 2 | Outcomes over 20 years (2015–2035) in people with HIV (age 15–65), according to monitoring strategy.

	No monitoring	Clinical monitoring	Clinical monitoring viral load confirmation	Clinical monitoring CD4 count confirmation	CD4 count monitoring (WHO)	CD4 count monitoring (<200)	Viral-load-informed differentiated care using DBS
Percentage of ART-experienced people who have fulfilled criterion for failure of first-line ART	7%	14%	10%	13%	41%	26%	27%
Percentage of ART-experienced people who have started second-line ART	3%	13%	10%	13%	38%	24%	25%
Percentage of people on ART who have (true) viral load <1,000 cps ml^{-1} (mean; over 20-year time horizon)	76%	79%	78%	79%	85%	82%	86%
Death rate (per 100 person years) among people on ART	4.43	3.63	4.06	3.67	3.02	3.07	3.18
Death rate (per 100 person years) among people with HIV	5.45	4.91	5.2	4.93	4.36	4.43	4.47
Death rate (per 100 person years) in the whole adult population	1.69	1.63	1.66	1.63	1.56	1.58	1.57
Death rate (per 100 person years) among people on ART who have virologically failed first-line ART (regardless of whether monitoring strategy has detected it)	9.94	7.5	8.66	7.62	5.53	5.79	5.85
Incidence of HIV (per 100 person years)	0.84	0.81	0.83	0.81	0.76	0.79	0.73

For each model run for each strategy, the outcome of interest (as listed in the first column) is output for each 3-month period between 2015–2035. Over 500 model runs are done for each strategy, then means are taken over 3-month periods and model runs. ART, antiretroviral therapy; Cps, copies; DBS, dried blood spot; WHO, World Health Organization.

In the base case we have considered there to be a switch rate of 0.5 per 3 months after the strategy-specific failure criteria have been met. In practice, in most settings, despite CD4 counts being measured, switching rates are much lower than this. We compared use of the CD4 count monitoring (WHO) strategy with a low switch rate of 0.05 per 3 months (the current situation in many countries) with viral-load-informed differentiated care with a switch rate of 0.5 per 3 months (Fig. 5). The results suggest that introduction of the viral-load-informed differentiated care using DBS accompanied by a high switch rate would lead to a substantial improvement in DALYs averted with a potential reduction in cost, compared with the current situation. In the simulated model population of Zimbabwe, over 20 years the CD4 count monitoring (WHO) strategy averts 540,000 DALYs compared with no monitoring at a cost of \$500 million, whereas viral-load-informed differentiated care using DBS averts 1.12 million DALYs compared with no monitoring at a cost of \$361 million.

We also consider only the viral-load-informed differentiated care strategy and assess the effect of variations in various aspects (Fig. 6); whether whole blood or plasma is used, whether the test is POC (central laboratory testing using whole blood is our DBS scenario above), the threshold to define failure (200, 1,000 or 5,000 cps ml^{-1}, which is only assessed in the context of plasma), and the frequency of measurement (every 6 months, annually or every 2 years). Monitoring every 6 months instead of annually averts more DALYs, but does not seem to be cost-effective at the \$500 threshold (ICER = \$1,234). Less frequent monitoring (such as every 2 years) would be cost-effective if it were to avert a similar number of DALYs to monitoring every year. However, implementing differentiated care based on viral-load monitoring as infrequently as every 2 years is currently untested and the potential health consequences are unknown, so this strategy is excluded from the comparison (Fig. 6a). Using the 5,000 cps ml^{-1} threshold also averts DALYs at a similar ICER to the 1,000 cps ml^{-1} threshold, but with reduced total benefit. Use of a whole blood sample (for example, DBS) instead of a plasma sample is not predicted to result in a marked difference in cost incurred (assuming the same unit cost per test) and a modest (4%) benefit in DALYs averted. There is a small (6%) predicted benefit of POC testing over laboratory monitoring in DALYs averted owing to the fact that the 3-month delay is avoided.

DISCUSSION

Our results suggest that viral-load-informed differentiated ART care, using DBS sampling if necessary, is likely to be cost-effective in low-income settings in sub-Saharan Africa and is a sustainable model for providing ART. That said, the level of savings that result from reduced clinic visits and that can be realized in practice with differentiated care are, so far, uncertain and require monitoring. The level of savings required depends partially on the cost of viral-load testing. With the viral-load test cost of \$22 as used in our base case, an annual saving of at least \$30 per year in those with viral suppression is required for viral-load-informed differentiated ART care to be cost-effective. Given that annual non-ART-programme costs average around \$80 per year[2] if patients are being seen every 1 to 3 months, a reduction in visit frequency to once every 6 months, and perhaps for long-term suppressed patients to every 9 to 12 months, should enable such savings. There is little evidence that patients seen at sites with higher non-ART-programme costs have better outcomes[2]. We estimate, based on modelling of Zimbabwe over 20 years, that in contrast to the current situation in many countries (CD4 count monitoring with low switch rates), introduction of viral-load-informed differentiated care would more than double the number of DALYs averted compared with no monitoring (1.12 million compared with 0.54 million) and deliver these at reduced costs (\$360 million compared with \$500 million).

A reduction in the frequency of clinic visits could also affect patients' adherence to ART and retention in care. There is evidence that some patients default from care because they are unable to keep up with the intensive clinic visit schedule owing to travel time and cost, and loss of work time[4].

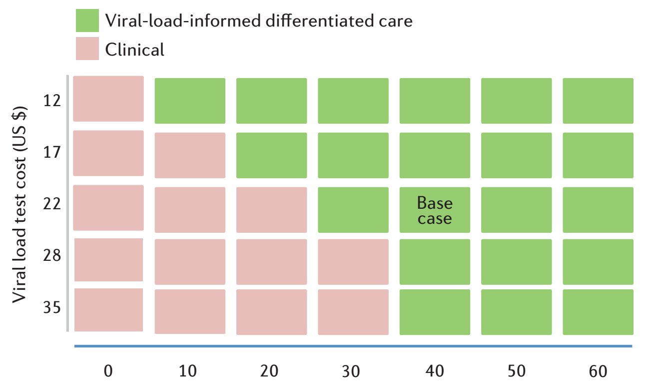


Figure 3 | Viral load cost-effectiveness. Indication of whether viral-load-informed differentiated care is the most cost-effective monitoring strategy according to cost of viral load tests and reduction in non-antiretroviral (ART) programme costs in people with viral suppression. In the context of cost-effectiveness threshold US$500. Colours indicate which monitoring strategy is economically preferred.

Notably, retention in care was more than 90% at 4 years among individuals enrolled in community ART clubs in Mozambique, owing, in part, to community-based adherence support, decreased travel requirements and patient preference[46,47]. We did not include in our model any such adherence or retention benefits associated with differentiated care. There is also the possibility that patients may feel less connected to care with a differentiated care model, and this might have adverse consequences for adherence and retention, although so far there is little evidence to suggest this.

When using the CD4 count to monitor people on ART, the WHO-recommended approach has been to define failure by a CD4 count of <100 cells mm^{-3} of blood or a decline from pre-ART baseline. Our modelling suggests that, given the high variability in CD4 count and the fact that it is not uncommon for people to interrupt ART for periods of time, this latter component results in low specificity and in many patients with viral suppression being incorrectly categorized as failing and hence switched unnecessarily. The alternative approach we evaluated, similar to that used in the DART trial[48], is to define failure based on a CD4 count of <100 mm^{-3} in years 1–3 on ART, and a CD4 count <200 mm^{-3} thereafter. This approach performed well in our modelling in terms of the death rate of people on ART (as it did in the trial itself), although it still resulted in a lower rate of viral suppression and hence a higher HIV incidence than with viral-load monitoring, resulting in poorer overall effectiveness. In settings that continue to have a CD4 count capacity, but not viral-load capacity, this suggests that the CD4 count monitoring (<200) strategy should be used until viral-load-informed differentiated care is introduced.

The requirement for frequent clinic visits is partially driven by shortages of ART supplies at a national level, resulting in clinic level rationing of ART quantities dispensed to patients at each visit. Increasing country buffer stocks, as well as improving forecasting of need, could enable longer drug supplies to be prescribed. However, even if it is not possible to prescribe more than a 1–2 months supply of drugs, various approaches can be considered to prevent patients from having to make frequent pharmacy-only visits to clinic[46,47,49–54]. These include community ART groups, whereby one person picks up the drugs for all the members or situations in which patients can pick up medicines in a shop or other non-clinical setting[55]. Other hurdles to overcome in adopting viral-load-driven reductions in frequency of clinical visits include obtaining buy-in from Ministries of Health for any required task shifting, and provision of human resources for dedicated adherence support for people with high viral load. In addition, support from professional associations of clinical, nursing and pharmacy staff will be important.

The fact that the viral load is a direct measure of the ongoing effect of treatment means it provides an ideal means to differentiate care provision. However, given the wider availability of CD4 count tests, it might be

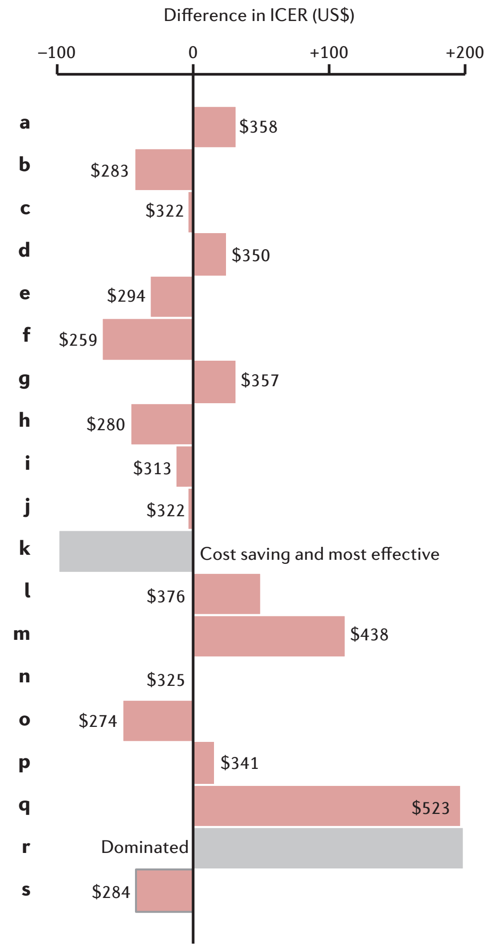


Figure 4 | Incremental cost-effectiveness ratio (ICER) for viral-load-informed differentiated care using dried blood spot (DBS) (compared with next less effective strategy on the efficiency frontier) according to changes in assumptions (see Supplementary Figure 1). **a,** DBS sensitivity 96% and specificity 79% for 1,000 copies per millilitre threshold (versus plasma). **b,** DBS sensitivity 71% and specificity 97% for 1,000 cps ml^{-1} threshold (versus plasma). **c,** DBS sensitivity 88% and specificity 93% for 1,000 cps ml^{-1} threshold (versus plasma). **d,** DBS sensitivity 85% and specificity 79% for 1,000 cps ml^{-1} threshold (versus plasma). **e,** Poorer population antiretroviral therapy (ART) adherence profile such that proportion with viral suppression with no monitoring/no second-line ART is 68% compared with 76% in base case and HIV incidence is 0.96 per 100 person years compared with 0.84 in base case. **f,** Future greater increase in sexual behaviour in population such that HIV incidence is 1.46 per 100 person years compared with 0.84 in base case. **g,** Permanent increase in adherence as a result of viral load measurement alert in none rather than 40%. **h,** Permanent increase in adherence as a result of viral load measurement alert in 100% rather than 40%. **i,** Policy of initiation of ART at diagnosis. **j,** Reduced frequency of visits if CD4 count measured >350 in past year. **k,** Switch rate of 0 (so only benefit of monitoring is to inform who should be seen less frequently). **l,** Lower prevalence of HIV in 2014 (6% instead of 15% in base case). **m,** Higher prevalence of HIV in 2015 (33% instead of 15% in base case). **n,** Lower proportion on ART in 2015 (33% instead of 56% in base case). **o,** Higher proportion on ART in 2015 (70% instead of 56% in base case). **p,** 5% discount rate instead of 3%. **q,** Ten year time horizon instead of 20 year. **r,** Two times higher rate of ART interruption if visit frequency has been reduced due to viral load being <1,000 cps ml^{-1}. **s,** Two times lower rate of ART interruption. Based on 200 model runs per strategy for each of **a**–**s**.

suggested that the CD4 count could be used instead. For example, visit frequency for people with a CD4 count of more than 350 mm^{-3} could be reduced. This would result in a similar reduction in clinic visit costs to

viral-load-informed differentiated care. The effectiveness of such an approach is unknown, however, and it would lead to some people in whom adherence is low and/or resistance is present, and viral load is high, being asked to visit clinic less frequently. It is well established that CD4 counts can remain high when virological failure is occurring[56] and, likewise, that the CD4 count can remain low despite full virologic suppression. Thus, the negative effects of such a strategy would be a concern and, although we did model this as a potential strategy (Supplementary Fig. 1j) it is possible that we did not fully capture the extent of these negative effects.

We have largely focussed on use of DBS rather than plasma collection as an approach. Although plasma samples from a venepuncture and sample separation are an ideal sample, for transport of more than 6–24 hours this requires cold temperatures and so the approach is only likely to be applicable in areas for which samples can reach the laboratory in that time.

Although we have argued that a DBS approach is feasible in most settings, this is not to say that the approach is working well everywhere[57]. It is important that there is investment in improvements to existing systems, including diagnostics laboratories and logistics of specimen distribution, and we have endeavoured to capture these costs as part of our overall costs of delivering viral-load testing using DBS. It is notable that most studies that have evaluated viral load using DBS compared with plasma have been performed in a laboratory setting using venepuncture samples and a capillary tube (which measures a precise 100 µl whole blood) to fill in the DBS card. Few studies are available to assess the performance of DBS in the real-world scenario — where it is hot, where sample-transport times are long, where venepuncture is not an option, and where samples are from a finger prick rather than a capillary tube — although one such study has found encouraging findings[27]. Our finding that viral-load-informed differentiated care is cost-effective was robust to low levels of sensitivity or specificity using DBS (Fig. 4, Supplementary Fig. 1).

We simplified the comparison of types of viral-load test by breaking them down according to whether they are done at POC or in a laboratory and whether the sample consisted of whole blood or plasma. We recognise that this is something of an oversimplification in that, for example, measurement of viral load by POC testing on whole blood may not always have the same sensitivity or specificity as using whole blood in the form of DBS. Improved sensitivity and specificity compared with DBS offers a modest, but real benefit, as does the ability to measure the viral-load level such that it can be acted on in the same day, avoiding a delay until the next visit or the need to contact and recall the patient. Even if a POC viral-load test with the desirable properties we considered does become available it is likely that countries would use a mix of approaches (plasma samples, DBS and POC) depending on settings. It should be noted that the cost we assumed for a POC assay of $22 was used as a placeholder for the actual cost when this is known. It is uncertain whether such tests will be able to be delivered at this cost, as a fully-loaded cost, which takes account of staff operator time, and our results should be interpreted in the light of this.

If differentiated care can be successfully implemented using viral-load monitoring less frequently than every 12 months (for example, every 24 months) our modelling suggests that less frequent monitoring would be expected to be cost-effective. However, the health risks of differentiated care with such infrequent viral-load monitoring are not well understood and may not have been fully captured in our model. Further evidence on whether this approach is feasible, and the health consequences of its implementation, is required. Only in highly resourced health-care systems (with a cost-effectiveness threshold of more than $1,400 per DALY averted) is more-frequent monitoring (for example, every 6 months) expected to be cost-effective.

We found little evidence to support substantial benefits associated with increasing or decreasing the cut-off (viral-load counts >1,000 cps ml^{-1}) at which treatment is considered to have failed. A cut-off of 200 cps ml^{-1} results in more DALYs being averted — due to identifying people with virological failure earlier — but relies on a plasma-based test (and phlebotomy to achieve sufficient sample volume) and does not meet the $500 cost-effectiveness threshold.

Given the role of viral-load testing for enabling reduced visit frequency, it should also have a role in people on second-line regimens. When evaluating

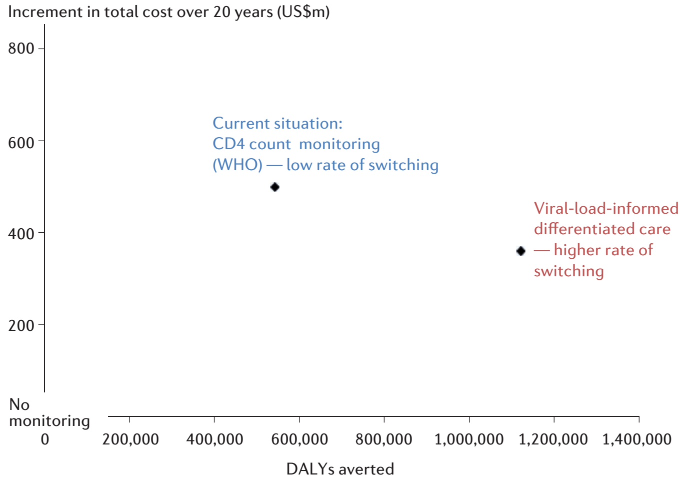


Figure 5 | Comparison with the current situation. The current situation, CD4 count (World Health Organization, WHO) monitoring with a low rate of switching in those meeting the failure criteria (0.05 per 3 months), and viral-load-informed differentiated care with switch rate as in our base case (0.5 per 3 months). DBS, dried blood spot.

our monitoring strategies we assumed that CD4 count or viral-load tests would only be done in patients on first-line treatment, so we may have understated the benefits of viral-load-informed differentiated care.

We considered whether our base case results would still hold with various alterations in assumptions and settings. In a scenario in which the pattern of adherence was generally poorer than in our base case (leading to 68% of people on ART with viral suppression compared with 82% in our base case) viral-load-informed differentiated care remained cost-effective. This was also true in a scenario with high HIV incidence rate, and scenarios with different HIV prevalence and ART coverage, suggesting that our findings should hold in various settings in the region.

Randomized trials have been performed to compare outcomes from CD4 count and viral-load monitoring, and these have not identified significant differences in outcomes. Such trials have been characterized by relatively short follow-up and low implementation of switching to second-line therapy[58–64], leading to low power to detect differences.

We focussed on monitoring for adults. In children and, more likely, adolescents levels of adherence may be lower than in adults. We did find that our main findings hold in populations with a tendency for lower adherence. However, there may be a greater reluctance to reduce visit frequency as children are growing up and constantly facing new challenges and situations and clinic staff may wish for regular contact to ensure that these new challenges have not led to a drop in adherence. Likewise, there may be a reluctance to reduce visit frequency for women in the year or so post-partum. We also considered whether monitoring more-intensively — every 6 months rather than every 12 months — would be cost-effective for populations with a poorer adherence profile (Supplementary Fig. 1t), but this was not the case. Other limitations of this work include the fact that we considered a hypothetical cohort with simulated outcomes, and future trends are uncertain — particularly in sexual behaviour, levels of male circumcision and adherence to ART. Furthermore, we assume continuation of HIV testing and ART availability at current trends. The profile of new POC viral-load tests is as yet uncertain, as is their cost. However, new diagnostic technologies, including POC viral-load testing and beyond, have great potential to enhance delivery of HIV care. We have investigated uncertainty through a series of one-way and multi-way sensitivity analyses and recognize that there are other approaches, such as probabilistic sensitivity analysis and approximate Bayesian computation that we intend to pursue in further work.

This work provides insight into how to deliver ART monitoring so that it is both effective and cost-effective. As well as providing some specific guidance to programmes, it highlights the need to research this area further to enable us to continue to understand the attributes of programmes and

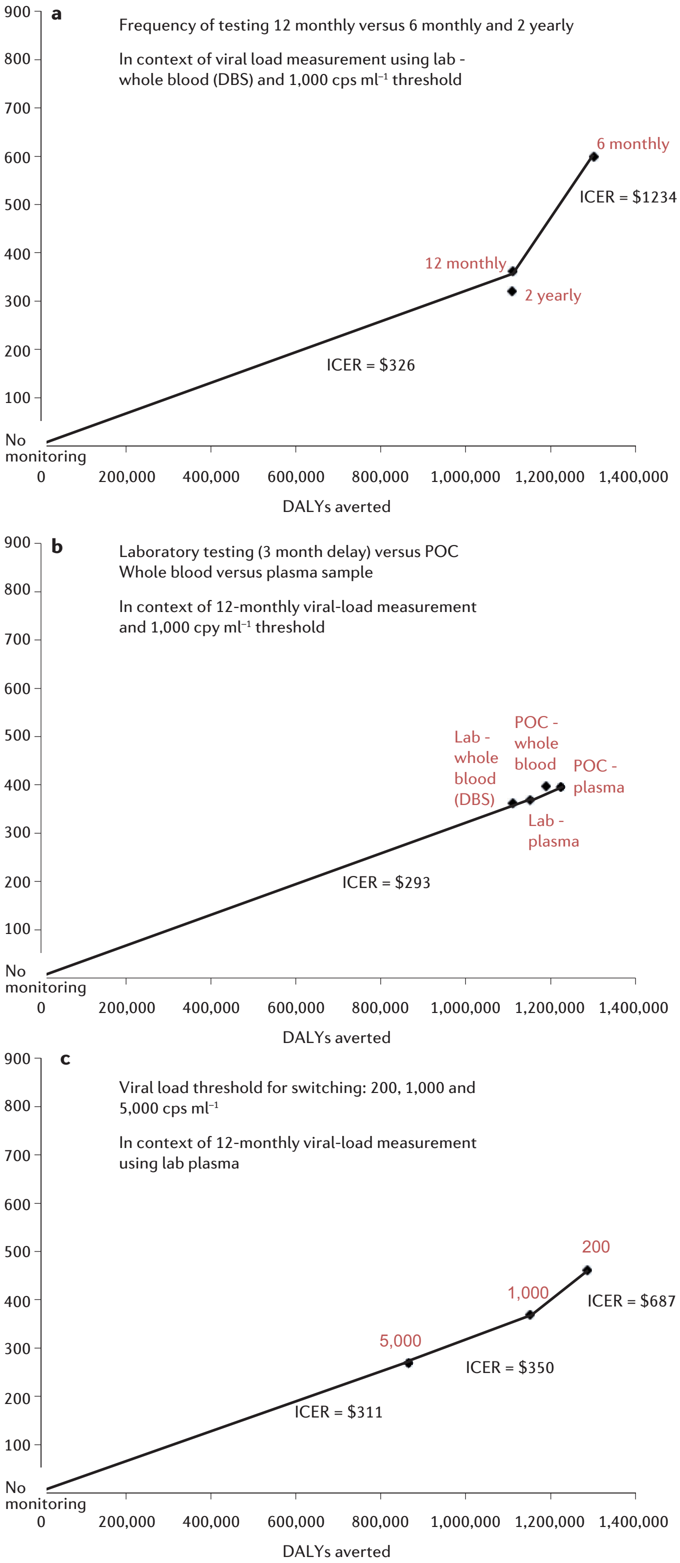


Figure 6 | Cost-effectiveness planes showing the effect of viral load measurement frequency, format and threshold, all in the context of viral-load-informed differentiated care. **a**, Viral load monitoring every 12-months is compared with every 6 months (every 2-year monitoring is excluded from the cost-effectiveness frontier due to unproven ability to base differentiated care on a 2-yearly value; however, if less frequent monitoring could be implemented without adverse health outcomes this would be cost-effective). **b**, Laboratory whole blood corresponds to dried blood spot (DBS). **c**, Alternative thresholds to define failure (viral load >200, >1,000 and >5,000 cps ml[1]) are compared in the context of laboratory monitoring every 12 months using plasma.

to determine how maximum health gains can be realized for patients with the resources available. We find that evidence is sufficient to recommend viral-load-informed differentiated care that uses DBS, but that further empirical confirmation as the approach is rolled out would be valuable.

1. Siapka, M., Remme, M., Dayo Obure, C., Maier, C., Dehne, K. L. & Vassall A. Is there scope for cost savings and efficiency gains in HIV services? A systematic review of the evidence from low- and middle-income countries. *Bull. World Health Organ.* **92,** 499–511 (2014).
2. Tagar, E. *et al.* Multi-country analysis of treatment costs for HIV/AIDS (MATCH): facility-level ART unit cost analysis in Ethiopia, Malawi, Rwanda, South Africa and Zambia. *PLoS ONE* **9,** e108304 (2014).
3. Menzies, N. A. *et al.* The cost of providing comprehensive HIV treatment in PEPFAR-supported programs. *AIDS* **25,** 1753–1760 (2011).
4. Ware, N. C. *et al.* Toward an understanding of disengagement from HIV treatment and care in sub-Saharan Africa: a qualitative study. *PLoS Med.* **10,** e1001369 (2013).
5. World Health Organisation. *Technical and Operational Considerations for Implementing HIV Viral Load Testing. Access to HIV diagnostics* (WHO, 2014).
6. UNAIDS. *Landmark HIV Diagnostic Access Program will Save $150m and Help Achieve New Global Goals on HIV* http://www.unaids.org/en/resources/presscentre/pressreleaseandstatementarchive/2014/september/20140925prviralload (UNAIDS, 2014)
7. Phillips, A. N., Pillay, D., Miners, A. H., Bennett, D. E, Gilks, C. F. & Lundgren, J. D. Outcomes from monitoring of patients on antiretroviral therapy in resource-limited settings with viral load, CD4 cell count, or clinical observation alone: a computer simulation model. *Lancet* **371,** 1443–1451 (2008).
8. Estill, J. *et al.* Viral load monitoring of antiretroviral therapy, cohort viral load and HIV transmission in Southern Africa: a mathematical modelling analysis. *AIDS* **26,** 1403–1413 (2012).
9. Braithwaite, R.S. *et al.* Alternative antiretroviral monitoring strategies for HIV-infected patients in east Africa: opportunities to save more lives? *J. Intl AIDS Soc.* **14,** 38 (2011).
10. Kimmel, A.D. *et al.* Laboratory monitoring to guide switching antiretroviral therapy in resource-limited settings: clinical benefits and cost-effectiveness. *J. Acquir. Immune Defic. Syndr.* **54,** 258–268 (2010).
11. Braithwaite, R. S. *et al.* How do different eligibility guidelines for antiretroviral therapy affect the cost-effectiveness of routine viral load testing in sub-Saharan Africa? AIDS **28** (Suppl 1)**,** S73–S83 (2013).
12. Estill, J. *et al.* Monitoring of antiretroviral therapy programmes in Malawi, South Africa and Zambia: Mathematical Modelling Study. *PLoS ONE* **8,** e57611.15 (2013).
13. Keebler, D. *et al.* Cost-effectiveness of different strategies to monitor adults on antiretroviral treatment: a combined analysis of three mathematical models. *Lancet Glob. Health* **2,** e35–e43 (2013).
14. Bendavid, E. *et al.* Cost-effectiveness of HIV monitoring strategies in resource-limited settings: a southern African analysis. *Arch. Int. Med.* **17,** 168 (2008).
15. Hamers, R. L., Sawyer, A. W., Tuohy, M., Stevens, W.S., Rinke de Wit, T. F & Hill, A. M. Cost-effectiveness of laboratory monitoring for management of HIV treatment in sub-Saharan Africa: a model-based analysis. *AIDS* **26,** 1663–1672 (2012).
16. Rutstein, S. E. et al. Dried blood spots for viral load monitoring in Malawi: feasible and effective. *PLoS ONE* **10,** e0124748 (2015).
17. Andreotti, M. *et al.* Correlation between HIV-1 viral load quantification in plasma, dried blood spots, and dried plasma spots using the Roche COBAS Taqman assay. *J. Clin. Virol.* **47,** 4–7 (2010).
18. Garrido, C. *et al.* Correlation between human immunodeficiency virus type 1 (HIV-1) RNA measurements obtained with dried blood spots and those obtained with plasma by use of Nuclisens EasyQ HIV-1 and Abbott RealTime HIV load tests. *J. Clin. Microbiol.* **47,** 1031–1036 (2009).
19. Johannessen, A. *et al.* Dried blood spots perform well in viral load monitoring of patients who receive antiretroviral treatment in rural Tanzania. *Clin. Infect. Dis.* **49,** 976–981 (2009).
20. Marconi, A. *et al.* Evaluation of the Abbott Real-Time HIV-1 quantitative assay with dried blood spot specimens. *Clin. Microbiol. Infect.* **15,** 93–97 (2009).
21. Arredondo, M. *et al.* Comparison of HIV-1 RNA measurements using plasma and dried blood spots (DBS) in the Automated Abbott Real Time Viral Load Assay. *J. Clin. Microbiol.* **50,** 569–572 (2011).
22. Pirillo MF-Pinson, P. *et al.* Quantification of HIV-RNA from dried blood spots using Siemens VERSANT® HIV-1 RNA (kPCR) assay. *J. Antimicrob. Chemother.* **66,** 2823–2826 (2011).
23. Ondoa, P. *et al.* Performance and logistical challenges of alternative HIV-1 virological monitoring options in a clinical setting of Harare, Zimbabwe. *BioMed Res. Int.* **2014,** 102598 (2014).
24. Fajardo, E. *et al.* Prospective evaluation of diagnostic accuracy of dried blood spots from finger prick samples for determination of HIV-1 Load with the NucliSENS Easy-Q HIV-1 version 2.0 Assay in Malawi. *J. Clinical Microbiol.* **52,** 1343–1351 (2014).
25. Smit, P. W. *et al.* Systematic review of the use of dried blood spots for monitoring HIV viral load and for early infant diagnosis. *PLoS ONE* **9,** e86461 (2014).
26. Rutstein, E. *et al.* Measures of viral load using Abbott RealTime HIV-1 Assay on venousand fingerstick dried blood spots from provider-collected specimens in Malawian District Hospitals. *J. Clin. Virol.* **60,** 392–398 (2014).
27. Mavedzenge, S. N. *et al.* Finger prick dried blood spots for HIV viral load measurement in field conditions in Zimbabwe. *PloS ONE* **10,** e0126878 (2015).

28. UNITAID. *HIV/AIDS Diagnostics Technology Landscape. Semi Annual Update*. www.unitaid.org (UNITAID, 2015).
29. World Bank. *Country and Lending Data* http://data.worldbank.org/about/country-and-lending-groups#Sub_Saharan_Africa (World Bank, 2015).
30. Duncombe, C. *et al*. Reframing HIV care: putting people at the centre of antiretroviral delivery. *Trop. Med. Int. Health* **20,** 430–447 (2015).
31. Roberts, T., Bygrave, H., Fajardo, E. & Ford, N. Challenges and opportunities for the implementation of virological testing in resource-limited settings. *J. Int. AIDS Soc*. **15,** 17324 (2012).
32. Cambiano, V. *et al*. Transmission of drug resistant HIV and its potential impact on mortality and treatment outcomes in resource-limited settings. *J. Infect. Dis*. **207,** S57–S62 (2013).
33. Cambiano, V. *et al*. Predicted levels of HIV drug resistance: potential impact of expanding diagnosis, retention, and eligibility criteria for antiretroviral therapy initiation. *AIDS* **28,** S15–S23 (2014).
34. Phillips, A. N. *et al*. Effectiveness and cost-effectiveness of potential responses to future high levels of transmitted HIV drug resistance in antiretroviral drug-naive populations beginning treatment: modelling study and economic analysis. *Lancet HIV* **1,** e85–e93 (2014).
35. Cambiano, V. *et al*. Assessment of the potential impact and cost-effectiveness of self-testing for HIV in low-income countries. *J. Infect. Dis*. http://dx.doi.org/*10.1093/infdis/jiv040* (2015).
36. Phillips, A. N. *et al*. Effect on transmission of HIV-1 resistance of timing of implementation of viral load monitoring to determine switches from first to second-line regimens in resource-limited settings. *AIDS* **25,** 843–850 (2011).
37. Madec, Y., Leroy, S., Rey-Cuille, M.-A., Huber. F & Calmy, A. Persistent difficulties in switching to second-line ART in sub-Saharan Africa — a systematic review and meta-analysis. *PLoS ONE* **8,** e82724 (2013)
38. Fox, M. P. *et al*. Rates and predictors of failure of first-line antiretroviral therapy and switch to second-line ART in South Africa. *J. Acquir. Immune Defic. Syndr*. **60,** 428–437 (2012).
39. Johnston, V., Fielding, K. L., Charalambous, S., Churchyard, G., Phillips, A. & Grant, A. D. Outcomes following virological failure and predictors of switching to second-line antiretroviral therapy in a South African treatment program. *J. Acquir. Immune Defic. Syndr*. **1,** 370–380 (2012).
40. Parkin, N. Measurement of HIV-1 viral load for drug resistance surveillance using dried blood spots: literature review and modelling of contribution of DNA and RNA. *AIDS Rev*. **16,** 160–171 (2014).
41. Claxton, K., Walker, S., Palmer, S. & Sculpher, M. *Appropriate Perspectives for Health Care Decisions*, *Centre for Health Economics Research Paper 54* (Univ. York, 2010).
42. Woods, E., Revill, P., Sculpher, M. & Claxton, K. *Country-Level Cost-Effectiveness Thresholds: Initial Estimates and the Need for Further Research*. https://www.york.ac.uk/media/che/documents/papers/researchpapers/CHERP109_cost-effectiveness_threshold_LMICs.pdf (Univ. York, 2015).
43. Salomon, J. A. *et al*. Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *Lancet* **380,** 2129–4310 (2012).
44. Hyle, E. P. *et al*. The clinical and economic impact of point-of-care CD4 testing in Mozambique and other resource-limited settings: a cost-effectiveness analysis. *PLoS Med*. **11,** e1001725 (2014).
45. MSF. *Untangling the Web of Antiretroviral Price Reductions* http://www.msfaccess.org/content/untangling-web-antiretroviral-price-reductions (MSF, 2014).
46. Decroo, T. *et al*. Four-year retention and risk factors for attrition among members of community ART groups in Tete, Mozambique. *Trop. Med. Int. Health*. **19,** 514–521 (2014).
47. Rasschaert, F. *et al*. Sustainability of a community-based anti-retroviral care delivery model — a qualitative research study in Tete, Mozambique. *J. Int. AIDS Soc*. **17,** 18910 (2014).
48. DART Trial Team. Routine versus clinically driven laboratory monitoring of HIV antiretroviral therapy in Africa (DART): a randomised non-inferiority trial. *Lancet* **375,** 123–131 (2010).
49. MSF. *ART Adherence Club Report and Toolkit* http://www.msf.org/sites/msf.org/files/cag_toolkit.pdf
50. MSF. *Community ART Group Toolkit. How to Implement the CAG Model. Bringing Treatment Closer to Home and Empowering Patients* http://samumsf.org/resources/toolkit-cag/ (MSF, 2015)
51. Simons, S. *et al. Ways to Reduce Costs and Improve Quality in Patient Monitoring: Barriers to Improvement, Buhera & Gutu District Experience, Zimbabwe. Consultation on Implementation Issues for Monitoring People on ART in Low-Income Settings in Sub-Saharan Africa* (MSF, 2015).
52. Babigumira, J. B. *et al*. Cost-effectiveness of a pharmacy-only refill program in a large urban HIV/AIDS clinic in Uganda. *PLoS ONE* **6,** e18193 (2011).
53. UNAIDS and MSF. *Community Based ART Delivery* (UNAIDS, 2015).
54. Grimsrud, A. *et al*. Implementation of community-based adherence clubs for stable antiretroviral therapy patients in Cape Town, South Africa. *J. Int. AIDS Soc*. **18,** 19984 (2015).
55. MSF and UNAIDS. *Closer to Home* http://www.msfaccess.org/sites/default/files/MSF_assets/HIV_AIDS/Docs/AIDS_report_ClosertoHome_ENG_2012.pdf (MSF, 2012).
56. Ledergerber, B. *et al*. Predictors of trend in CD4-positive T-cell count and mortality among HIV-1-infected individuals with virological failure to all three antiretroviral-drug classes. *Lancet* **364,** 51–62 (2004).
57. Terris-Prestholt F. First-line antiretroviral therapy for HIV-infected children. *AIDS* **29,** 1261–1262 (2015).
58. Mermin, J. *et al*. Utility of routine viral load, CD4 cell count, and clinical monitoring among adults with HIV receiving antiretroviral therapy in Uganda: randomised trial. *Br. Med. J*. **343,** d6792 (2011).
59. Kahn, J. G. *et al*. CD4 cell count and viral load monitoring in patients undergoing antiretroviral therapy in Uganda: cost effectivness study. *Br. Med. J*. **343,** d6884 (2011).
60. DART Trial Team. Routine versus clinically driven laboratory monitoring of HIV antiretroviral therapy in Africa (DART): a randomised non-inferiority trial. *Lancet* **375,** 123–131 (2010).
61. Jourdain, G. *et al*. Switching HIV treatment in adults based on CD4 count versus viral load monitoring: a randomized, non-inferiority trial in Thailand. *PLoS Med* **10,** e1001494 (2013).
62. Saag, M. *et al*. Cluster randomized trial of routine vs discretionary viral load monitoring among adults starting ART: Zambia. *19th Conf. Retroviruses and Opportunistic Infections* (2012).
63. Laurent, C. *et al*. Monitoring of HIV viral loads, CD4 cell counts, and clinical assessments versus clinical monitoring alone for antiretroviral therapy in rural district hospitals in Cameroon (Stratall ANRS 12110/ESTHER): a randomised non-inferiority trial. *Lancet Infect. Dis*. **11,** 825–833 (2011).
64. Boyer, S. *et al*. Monitoring of HIV viral load, CD4 cell count, and clinical assessment versus clinical monitoring alone for antiretroviral therapy in low-resource settings (Stratall ANRS 12110/ESTHER): a cost-effectiveness analysis. *Lancet Infect. Dis*. **13,** 577–586 (2013).

SUPPLEMENTARY MATERIAL

Is linked to the online version of this paper at: http://dx.doi.org/10.1038/nature16046

ACKNOWLEDGEMENTS

Funding for the study was received from the HIV Modelling Consortium, the Bill and Melinda Gates Foundation, the World Health Organization and the HIV Diagnostics Modelling Consortium. We are grateful for the use of cluster computing facilities at UCL, Legion@UCL, without which this work would not have been possible. J. L. is supported by the Danish National Research Foundation (DNRF:126).

APPENDIX

Andrew Phillips[1], Amir Shroufi[2], Lara Vojnov[3], Jennifer Cohn[4], Teri Roberts[4], Tom Ellman[2], Kimberly Bonner[5], Christine Rousseau[6], Geoff Garnett[6], Valentina Cambiano[1], Fumiyo Nakagawa[1], Deborah Ford[7], Loveleen Bansi-Matharu[1], Alec Miners[8], Jens D. Lundgren[9], Jeffrey W. Eaton[10], Rosalind Parkes-Ratanshi[11], Zachary Katz[3], David Maman[2], Nathan Ford[12], Marco Vitoria[12], Meg Doherty[12], David Dowdy[13], Brooke Nichols[14], Maurine Murtagh[15], Meghan Wareham[3], Kara M. Palamountain[16], Christine Chakanyuka Musanhu[17], Wendy Stevens[18], David Katzenstein[19], Andrea Ciaranello[20], Ruanne Barnabas[21], R. Scott Braithwaite[22], Eran Bendavid[23], Kusum J. Nathoo[24], David van de Vijver[14], David P. Wilson[25], Charles Holmes[26], Anna Bershteyn[27], Simon Walker[28], Elliot Raizes[29], Ilesh Jani[30], Lisa J. Nelson[31], Rosanna Peeling[32], Fern Terris-Prestholt[33], Joseph Murungu[34], Tsitsi Mutasa-Apollo[34], Timothy B. Hallett[10] & Paul Revill[28].

[1]Department of Infection and Population Health, University College London, Rowland Hill Street, London NW3 2PF, UK. [2]Southern Africa Medical Unit (SAMU), Medecins sans Frontieres (MSF) SA, Waverley Business Park, Wyecroft Rd, Mowbray 7700, Cape Town, South Africa. [3]Clinton Health Access Initiative, 383 Dorchester Avenue, Boston, Massachusetts 02127, USA. [4]Médecins Sans Frontières, Access Campaign, rue du Lausanne 82, 1202 Geneva Switzerland. [5]Médecins Sans Frontières, 78 rue de Lausanne, Case Postale 116, 1211 Geneva 21, Switzerland. [6]Bill and Melinda Gates Foundation, PO Box 23350, Seattle, Washington 98199, USA. [7]MRC Clinical Trials Unit at UCL, Institute of Clinical Trials & Methodology, Aviation House, 125 Kingsway, London WC2B 6NH, UK. [8]Health Services Research & Policy, London School of Hygiene and Tropical Medicine, Room 134, 15-17 Tavistock Place, London WC1H 9SY, UK. [9]CHIP, Department of infectious diseases, Rigshospitalet, University of Copenhagen, Blegdamsvej 92100 Copenhagen, Denmark. [10]Department of Infectious Disease Epidemiology, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK. [11]Infectious Diseases Institute (IDI), College of Health Sciences, Makerere University, PO Box 22418, Kampala, Uganda. [12]HIV/AIDS and Global Hepatitis Programme, World Health Organization, 20 Ave Appia 1211, Geneva, Switzerland. [13]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street E6531, Baltimore, Maryland 21205, USA. [14]Department of Viroscience, Erasmus Medical Center, PO Box 20403000CA Rotterdam, the Netherlands. [15]International Diagnostics Centre, London School of Hygiene & Tropical, Medicine, Keppel Street, London WC1E 7HT, UK. [16]Kellogg School of Management, Northwestern University, 2001 Sheridan Road Evanston, Illinois 60208, USA. [17]WHO Country Office 86 Enterprise Road Cnr, Glenara PO Box CY 348, Causeway Harare, Zimbabwe. [18]Department of Molecular Medicine and Haematology, University of the Witwatersrand, South Africa. [19]Division of Infectious Disease, Laboratory Grant Building S-146, Office Lane 154, Stanford University Medical Center, 300 Pasteur Drive, Stanford, California 94305-5107, USA. [20]Massachusetts General Hospital Division of Infectious Diseases, 50 Staniford Street, 936 Boston, Massachusetts 02114, USA. [21]Medicine, Global Health and Epidemiology, University of Washington (UW), 325 9th Avenue, Seattle, Washington 98104, USA. [22]Department of Population Health, New York University School of Medicine, 227 East 30th Street Office 615, New York, New York 10016, USA. [23]Division of General Medical Disciplines, Department of Medicine Stanford University, MSOB 1265 Welch Road x332 Stanford,

California 94305, USA. [24]University of Zimbabwe, College of Health Sciences, Department of Paediatrics and Child Health, PO Box A178, Avondale, Harare, Zimbabwe. [25]University of New South Wales, Level 6, Wallace Wurth Building, UNSW Campus, Sydney, New South Wales 2052, Australia. [26]Centre for Infectious Disease Research in Zambia, 5032 Great North Road, Lusaka, Zambia. [27]Institute for Disease Modeling, 3150 139th Avenue SE, Bellevue, Washington 98005, USA. [28]Centre for Health Economics, University of York, Heslington, York YO10 5DD, UK. [29]Care and Treatment Branch Center for Global Health, Division of Global HIV/AIDS (GAP), CDC, MS-E04, 1600 Clifton Road NE, Atlanta, Georgia 30333, USA. [30]Instituto Nacional de Saúde (INS), Ministry of Health, PO Box 264, Maputo, Mozambique. [31]The Office of the US Global AIDS Coordinator and Health Diplomacy (S/GAC), U.S. Department of State, SA-22, Suite 10300, 2201 C Street, Washington DC 20520, USA. [32]Clinical Research Department, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, UK. [33]Department of Global Health and Development, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK. [34]Ministry of Health and Child Care, P. O CY 1122, Causeway, Harare, Zimbabwe.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests. Financial support for this publication has been provided by the Bill & Melinda Gates Foundation.

ADDITIONAL INFORMATION

ARTICLE OPEN

Systematic review and meta-analysis of community and facility-based HIV testing to address linkage to care gaps in sub-Saharan Africa

Monisha Sharma[1], Roger Ying[2], Gillian Tarr[1] & Ruanne Barnabas[1,2,3,4]

HIV testing and counselling is the first crucial step for linkage to HIV treatment and prevention. However, despite high HIV burden in sub-Saharan Africa, testing coverage is low, particularly among young adults and men. Community-based HIV testing and counselling (testing outside of health facilities) has the potential to reduce coverage gaps, but the relative impact of different modalities is not well assessed. We conducted a systematic review of HIV testing modalities, characterizing community (home, mobile, index, key populations, campaign, workplace and self-testing) and facility approaches by population reached, HIV positivity, CD4 count at diagnosis and linkage. Of 2,520 abstracts screened, 126 met eligibility criteria. Community HIV testing and counselling had high coverage and uptake and identified HIV-positive people at higher CD4 counts than facility testing. Mobile HIV testing reached the highest proportion of men of all modalities examined (50%, 95% confidence interval (CI) = 47–54%) and home with self-testing reached the highest proportion of young adults (66%, 95% CI = 65–67%). Few studies evaluated HIV testing for key populations (commercial sex workers and men who have sex with men), but these interventions yielded high HIV positivity (38%, 95% CI = 19–62%) combined with the highest proportion of first-time testers (78%, 95% CI = 63–88%), indicating service gaps. Community testing with facilitated linkage (for example, counsellor follow-up to support linkage) achieved high linkage to care (95%, 95% CI = 87–98%) and antiretroviral initiation (75%, 95% CI = 68–82%). Expanding home and mobile testing, self-testing and outreach to key populations with facilitated linkage can increase the proportion of men, young adults and high-risk individuals linked to HIV treatment and prevention, and decrease HIV burden.

Nature 528, S77–S85 (3 December 2015), DOI: 10.1038/nature16044
This article has not been written or reviewed by *Nature* editors. *Nature* accepts no responsibility for the accuracy of the information provided.

Globally, there are around 2.3 million new HIV infections annually, 80% of which occur in sub-Saharan Africa[1]. Despite the high burden, only one-third of adults in sub-Saharan Africa have been tested for HIV in the past year and less than 50% of HIV-positive individuals know their status[2,3]. Knowledge of one's serostatus is vital for accessing lifesaving antiretroviral therapy (ART) and linking to HIV prevention. Conventional facility-based HIV testing and counselling (HTC) has not achieved high testing coverage in sub-Saharan Africa and will probably be insufficient to meet UNAIDS ambitious 90-90-90 targets — 90% of HIV-positive people knowing their status, 90% of HIV-positive people who are aware of their status on ART, and 90% of people on ART virally suppressed[4,5]. Barriers to facility testing include distance from clinic, long wait times, costs (transportation, lost wages and childcare), confidentiality concerns, low perceived risk and infrequent contact with the health-care system[6]. In addition, patients often present at facilities late in the course of their illness, increasing HIV morbidity, mortality and transmission[7]. Community-based HTC (conducted outside of a health facility) has the potential to overcome these barriers, achieve high coverage, and identify asymptomatic HIV-positive individuals at high CD4 counts[8,9]. In addition, community HTC may reach more men, young adults, and key populations than facility HTC. Community-based strategies also require minimal infrastructure allowing for easier scale up[10–12].

Community HTC modalities include: home, mobile, workplace, index partner/family members (sexual partners or family members of HIV-positive individuals) and as part of a campaign. Uptake and demographics of populations reached can vary widely by modality[9]. A large number of studies on HTC have been conducted in sub-Saharan Africa and a previous systematic review was completed in 2012, but facility testing was not included and uptake in men and young adults was not assessed. In addition, several large-scale interventions have been published since 2012 (refs 11, 13–15). Recently, the World Health Organization released guidelines that strongly recommend implementing community HTC[16]. As most countries have multiple and varying epidemics, UNAIDS recommends creating regional policies tailored to the macroepidemic rather than nationwide approaches[17]. Local policymakers will need to determine the optimal combination of community HTC interventions to increase testing in the context of their country's HIV epidemic.

To provide evidence for decision makers, we summarize the literature on community and facility-based HTC. We characterize each modality by population coverage, since high coverage is beneficial to both HIV-positive and -negative people. HTC can reduce risk behaviour in HIV-negative individuals, while providing a means to link them to primary prevention (including circumcision and pre-exposure prophylaxis (PrEP))[18–21]. We evaluate effectiveness in reaching men and young adults (both groups have low HIV testing and poorer clinical

[1]Department of Epidemiology, University of Washington, 1959 NE Pacific Street, Seattle, Washington 98195, USA. [2]Department of Global Health, University of Washington, 1510 San Juan Road 310e, Seattle, Washington 98195, USA. [3]School of Medicine, University of Washington, 4333 Brooklyn Avenue NE, Seattle, Washington 98105, USA. [4]Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., Seattle, Washington 98109, USA. Correspondence should be addressed to: M. S. e-mail: msharma04@gmail.com.

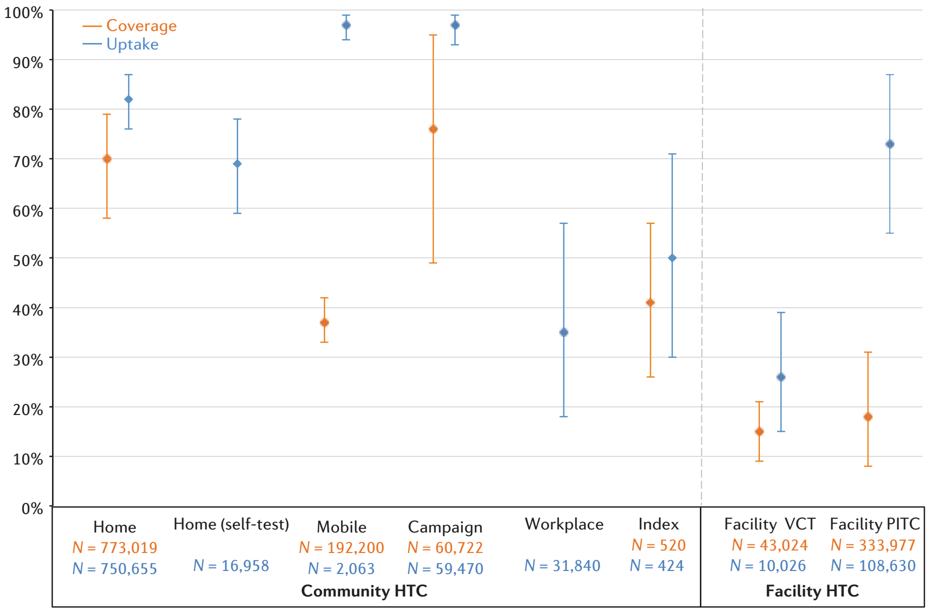


Figure 1 | Pooled coverage and uptake of HIV testing and counselling (HTC) modalities. Coverage is defined as total number of people tested/total number of people in the target population. Uptake is defined as total number of people tested/total number of people offered testing. Bars indicate 95% confidence intervals of random effects meta-analyses. *N*, sample size; PITC, provider-initiated testing and counselling; VCT, voluntary counselling and testing.

outcomes once infected[22–24]) and targeted HTC for key populations (men who have sex with men (MSM), commercial sex workers (CSWs) and people who inject drugs (PWID)) — groups that generally have very high HIV prevalence and low access to health care[25]. We assess HIV positivity to characterize yield and examine CD4 count at diagnosis to identify modalities that have the potential to link infected individuals to care earlier in their disease course. Estimates from our analysis can also be used as parameters in mathematical models to project the long-term impact of HTC interventions.

METHODS

Inclusion criteria. We conducted a systematic literature review following Cochrane and PRISMA (preferred reporting items for systematic reviews and meta-analyses) guidelines[26]. Studies were eligible for inclusion if they reported data on at least one of the following outcomes: coverage (individuals who accepted HTC/eligible target population); uptake (individuals who accepted HTC/individuals offered HTC); proportion of young adults (either under 25 or under 30 years); proportion of men; proportion of first-time testers; HIV positivity (number positive/total tested); proportion with a CD4 count of 350 cells μl^{-1} or less; proportion linked to care (those who had visited a clinic, obtained a CD4 count or initiated ART); proportion retained in care (individuals retained/individuals who initiated ART); or cost per person tested. The target population was defined as the eligible population in the catchment area, either enumerated by the study (often the case for home HTC) or estimated (often the case for mobile and campaign HTC). For facility HTC, the target population was defined as people visiting the clinic, and for index partner or family members it was defined as all sexual partners or cohabitating family members listed by the index patient. With the exception of HTC targeted to key populations, we excluded HTC studies not related to general population screening, including case reports and studies limited to antenatal or paediatric settings, or to patients with specific diseases (for example, tuberculosis). Observational (cross-sectional and cohort) studies and randomized trials were eligible for inclusion. Studies were included in the analyses more than once if they had different arms or multiple study sites (for example, urban and rural settings or different countries). If more than one wave of a survey or intervention was completed, only the most recent was used.

Search strategy. Literature searches were conducted with the help of a librarian on 22 July 2014 and updated on 10 June, 2015. Briefly, we searched PubMed, EMBASE, Cochrane Library, Global Health Database, African Index Medicus, and conference abstracts (CROI, R4P, IAS) using MeSH terms for PubMed and comparable terms for other databases. Search terms included "HIV infections/diagnosis" AND "Africa South of the Sahara" AND ("mass screening" OR test OR tests OR testing OR screen* OR diagnosis OR "counseling"). Bibliographies of relevant papers were screened and authors were contacted for missing outcomes. Searches were limited to human studies published between 2000 and 2015. The full strategy is described in the Supplementary Information.

Definitions of HTC modalities. Community-based HTC was defined as testing conducted outside of health facilities. Facility-based HTC was conducted in health-care facilities (clinics, hospitals, fixed stand-alone voluntary counselling and testing sites). Facility HTC was divided into two categories: voluntary counselling and testing (VCT), which is patient-initiated testing and provider-initiated testing and counselling (PITC), which is routine; or opt-out HTC that is initiated by a provider. Community HTC modalities included home (offering HTC door-to-door to a catchment area), mobile (setting up a mobile van or container to provide HTC in a central area of a community), index partner or family member (offering HTC to individuals who may have been exposed to HIV by a sexual partner or who have an HIV-positive household member), campaign (short — generally 1 to 2 weeks — intensive community mobilization followed by mobile testing, often partnered with other health interventions), key populations (targeted to MSM, CSWs and PWID) and workplace (offered at a place of employment). We examined a subset of home and workplace HTC that used self-testing.

Data screening and extraction. M.S., R.Y. and R.V.B. screened abstracts for initial inclusion. Disagreements were adjudicated by reviewing the full text. M.S., R.V.B., R.Y. and G.T. reviewed papers for eligibility and used a standardized extraction form to characterize eligible studies (Supplementary Information 2). Study quality was rated low, moderate or high based on representativeness of underlying population, follow-up (present or absent), assessment of outcomes, and number of outcomes presented. Costs were inflated to 2012 US dollars by converting to local currency units, multiplying by the ratio of each country's gross domestic product deflator (2012 deflator divided by base year deflator) and converting back to US dollars[27].

Statistical analysis. Random effects meta-analysis of single proportions with binomial exact confidence intervals (CI) was used to summarize results. Proportions were stabilized using the Freeman–Tukey double arcsine transformation unless the number of events was less than ten, in which case a logit

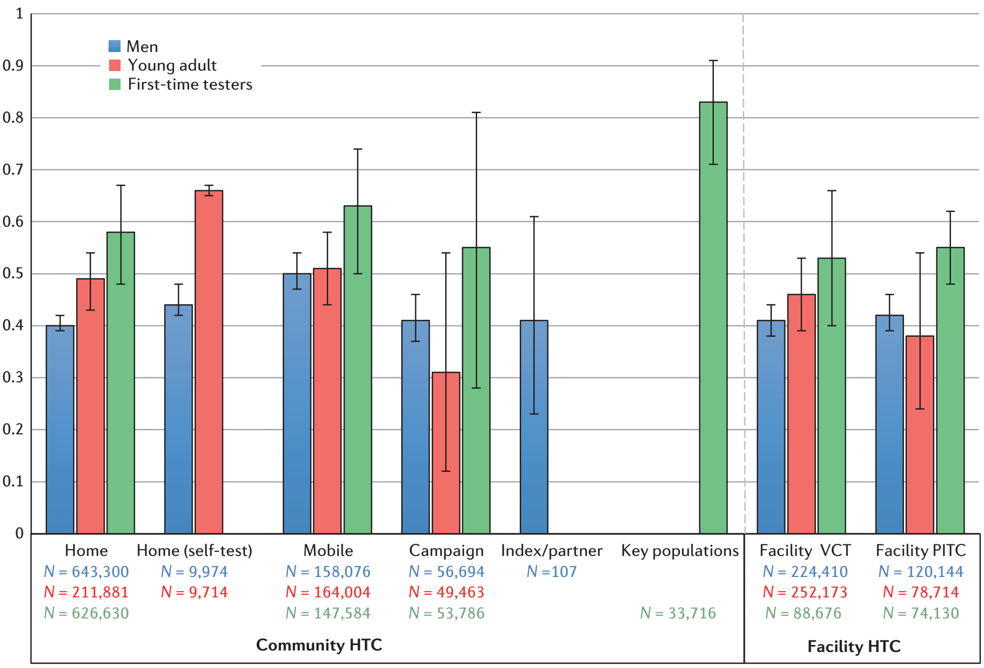


Figure 2 | Pooled percentage of men, young adults and first-time testers by HIV testing and counselling (HTC) modality. Bars indicate 95% confidence intervals of random effects meta-analyses. *N*, sample size. PITC, provider-initiated testing and counselling; VCT, voluntary counselling and testing.

transformation was used because of convergence issues. Heterogeneity was quantified using the I^2 statistic. For modalities with enough data (ten studies or more), trends were examined by year before 2005 (when the HIV rapid diagnostic test was introduced), country and facilitated linkage. Analyses were conducted in R software using the metaprop function in the meta package[28].

RESULTS

We identified 126 eligible studies out of 2,520 abstracts (Supplementary Figure S0.a). Overall, 64% of studies were rated moderate or high quality (Supplementary Information 2). Most studies included in our analysis evaluated facility and home HTC. We identified far fewer studies on other types of community HTC: home with self-testing (n = 2), workplace with self-testing (n = 2), index partner/family member (n = 5), key populations (n = 5), campaign (n = 5) and workplace (n = 4). Forest plots of each outcome by modality are provided in the Supplementary Information with pooled estimates presented here. I^2 values of pooled estimates varied from 90% to 100%, reflecting high heterogeneity in study designs and countries included (Supplementary Information). The countries represented varied by outcome with the greatest number of countries having data for home and facility HTC coverage, uptake and tester demographics. Far fewer studies reported CD4 count at diagnosis and linkage to care outcomes; studies containing these data were mainly conducted in South Africa, Kenya and Uganda. All home self-testing studies were conducted in Malawi and the most key population studies were conducted in Nigeria. Overall, the largest number of studies were conducted in South Africa.

Coverage and uptake

Coverage was reported in 19 home HTC studies[15,18,29-45], 1 mobile[13], 2 campaign[46,47], 3 index partner/family member[48-50], 5 facility VCT[51-55], and 5 facility PITC studies[56-61]. Overall, community HTC modalities achieved higher coverage than facility, with home (70%, 95% CI = 58–79) and campaign (76%, 95% CI = 49–95%) having the highest population coverage (Fig. 1). Home HTC consistently achieved high coverage across 19 studies, whereas campaign coverage was also high, but based on only two studies. Pooled coverage was 37% (95%, CI = 33–42%) for mobile HTC, from 1 study conducted in 3 countries (South Africa, Tanzania and Zimbabwe). Coverage of index HTC was heterogeneous depending on target group (family members or sexual partners) and type of contact tracing (active or passive referral). Figure 1 shows results for sexual partner tracing only (41%); full results are shown in Supplementary Figure S18. Facility VCT (15%, 95% CI = 9–21%) and PITC (18%, 95% CI = 18–31%) attained the lowest coverage.

Uptake was reported in 31 home HTC studies[5,14,15,18,27,29-38,40-45,53,62-74], 2 home with self-testing[11,75], 2 mobile[10,68], 3 index partner or family member[48-50], 4 campaign[46,47,76,77], 3 workplace[78-80], 3 facility VCT[54,56,81], and 11 facility PITC studies[56,57,59,60,81-87]. Overall, community modalities had high uptake (Fig. 1). Home HTC had a pooled uptake of 82% (95% CI = 76–87%) and home with self-testing had slightly lower uptake (69%, 95% CI = 59–78). Mobile and campaign had the highest uptake (both 97%). Index uptake was 89% (95% CI = 88–90%) for home testing of family members (Supplementary Figure S10) and 52% for sexual partners (95% CI = 30–71%; Fig. 1). Uptake for facility VCT was defined as number tested divided by number referred for VCT by provider, for facility PITC it was defined as number tested divided by number offered PITC. We found higher uptake for people given routine PITC (73%, 95% CI = 55–87%) compared with those referred to on site VCT (26%, 95% CI = 15–39%).

Demographics of testers

The percentage of men out of total persons tested was reported in 25 home HTC studies[5,14,18,29,31,32,37,38,41-45,63,64,66,68-72,88-90], 2 home with self-testing[11,75], 10 mobile[10,13,68,72,91-99], 3 index partner[47,49,88], 3 campaign[46,47,76], 2 workplace[100,101], 20 facility VCT[52,54,61,64,81,88,89,92,93,95,96,98,102-107], and 13 facility PITC[58,60,82-84,86,99,108-113] (Fig. 2). Mobile had the highest percentage of men (50%, 95% CI = 47–54%), whereas home had the lowest for general population HTC (40%, 95% CI = 39–41%). Index partner testing had 41% men (95% CI = 23–61%), but varied greatly by tracing strategy; active tracing had 50% men whereas passive clinic referral had only 15% (Supplementary Figure S18). Facility VCT and PITC both had 42% men.

Percentage of participants reporting testing for the first time was included in 20 home HTC studies[5,14,18,29,31,32,38,41-44,63,65,66,68-72,88], 11 mobile[10,12,68,93-95,97,103,114], 3 campaign[46,47,77], 3 key populations[25,115,116], 7 facility VCT[12,54,91,93,95,106], and 5 facility PITC[58,86,88,111,112]. Pooled percentages of first-time testers were higher for community than facility modalities (Fig. 2). Percentages varied by country, with South Africa consistently having the lowest percentage of first-time testers across modalities (Supplementary Figures S23–S27). Key population interventions had the highest proportion of first-time testers (83%, 95% CI = 71–91%), and mobile had the highest percentage among the general population (63%, 95% CI = 50–74%). Home HTC had 58% first-time testers (95% CI = 48–67%), and campaign had 55% (95% CI = 20–91%), but was highly variable depending on the setting (Supplementary Figure S25). Facility VCT had 53% (95% CI =

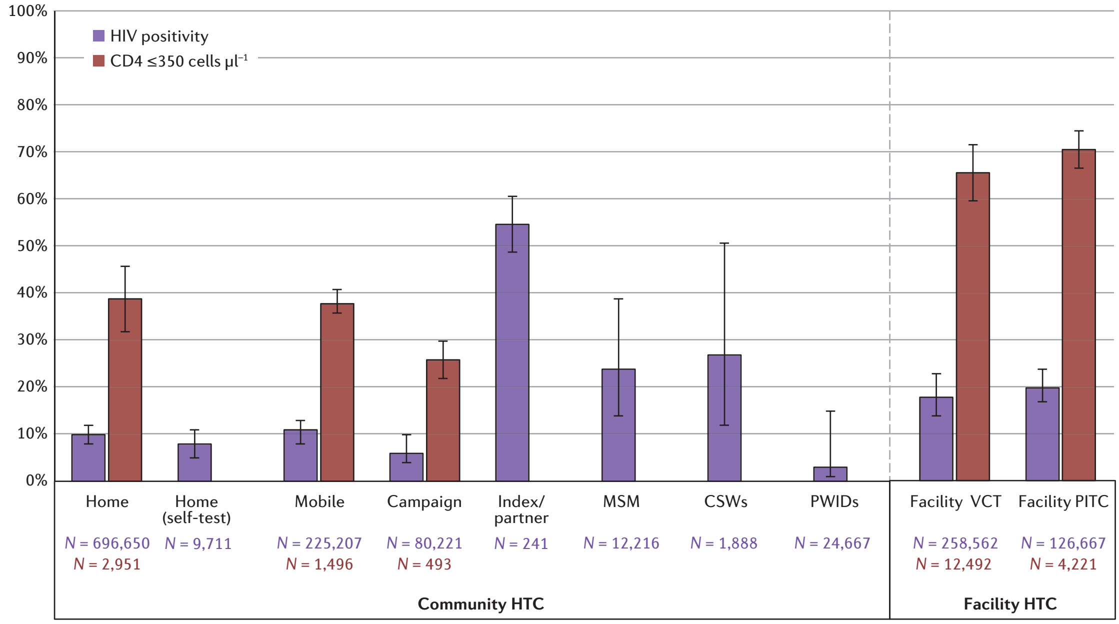


Figure 3 | Pooled HIV positivity and proportion of newly diagnosed HIV positivity with CD4 count of 350 cells μl^{-1} or less by HIV testing and counselling (HTC) modality. Bars indicate 95% confidence intervals of random effects meta-analyses. *N*, sample size. CSWs, commercial sex workers; MSM, men who have sex with men; PITC, provider-initiated testing and counselling; PWID, people who inject drugs; VCT, voluntary counselling and testing.

40–66%) and PITC had 55% (95% CI = 48–62%) first-time testers.

The percentage of young adults testers (either under 25 or 30 years) was reported in 17 home HTC studies[5,18,29–31,35,37,38,45,63,64,68–70,73,74,90,117], 1 home with self-testing[11], 13 mobile[10,12,13,68,91,93,95–97,103,107,114], 2 index partner[48,88], 2 campaign[47,77], 20 facility VCT[12,51,52,54,64,88,89,91–93,95,104–107,114,118–120], and 6 facility PITC[58,82,86,88,110,113]. Results varied considerably by study (Supplementary Figures S29–S35). Community HTC generally tested a higher proportion of young adults than facility modalities; home with self-testing had the largest percentage (66%, 95% CI = 65–67%), followed by mobile, and then home (Fig. 2). Campaign reported 31% of young adults, but varied from 20–50% depending on the study (Supplementary Figure S32). Facility VCT had 46% (95% CI = 39–53%) and PITC had 38% (95% CI = 39–53%).

HIV positivity and CD4 count ≤350 cells ml^{-1}

Yield of HIV-positive people (HIV positivity) was reported in 29 home studies[14,15,18,27,29–32,34,36,38,41–45,63,65,66,68,70–73,88,89], 1 home with self-testing[11], 12 mobile[10,13,68,72,92–95,97,98,103,107,114], 5 campaign[46,47,76,77,120], 3 workplace[79,80,121], 4 key population[12,115,116,122], 4 index partner[48–50,88], 27 facility VCT[54–56,64,81,84,88,91–93,95,98,102,104–107,114,118–120,123–127], and 17 facility PITC[56,57,59,60,81,83–88,99,110–113,126] studies. Community-based strategies for the general population had lower HIV positivity (6–11%) than facility HTC (18–20%), whereas targeted community HTC for key populations and sexual partners of index patients had the highest HIV yield (Fig. 3). HTC interventions targeting sexual partners of index cases had 55% positivity (95% CI = 49–61%), those for MSM had 24% (95% CI = 14–39%), for CSWs had 27% (95% CI = 12–51%), and interventions targeting PWIDs had the lowest positivity of 3% (95% CI = 1–15%). Index HTC for family members had similar HIV yield to home and mobile HTC (9%, 95% CI = 5–14%) (Supplementary Figure S42). Forest plots of HIV positivity for each modality stratified by country are shown in Supplementary Figures S36–S44). HIV positivity for community HTC in the general population largely mirrored prevalence of the country where the study was conducted, with the exception of four countries with the highest prevalence: Mozambique, Swaziland, Botswana and Lesotho. These countries have adult HIV prevalence ranging from 22 to 27% (ref. 128), but HIV yield from home, mobile and campaign HTC was 5–12%. HIV positivity for facility VCT and PITC was generally higher than prevalence in the general population.

The proportion of individuals with a CD4 count of 350 cells μl^{-1} or less at HIV diagnosis was reported in 7 home[14,38,42,43,65,72,73], 3 mobile[91,94,114], 3 campaign[46,47,76], 8 facility VCT[60,81,107,126,127,129–131] and 5 facility PITC studies[61,81,99,126,130]. Community-based strategies identified HIV-positive individuals at higher CD4 counts than facility HTC, with campaign having the lowest proportion with a CD4 count of 350 cells μl^{-1} or less (26%, 95% CI = 22–30%) (Fig. 3). Home (39%, 95% CI = 32–46%) and mobile (38%, 95% CI = 36–41%) had similar proportions of HIV-positive individuals with a CD4 count of 350 cells μl^{-1} or less, whereas facility VCT (66%, 95% CI = 60–72%) and PITC (71%, 95% CI = 67–75%) had the highest proportion.

Linkage and retention in care for HIV-positive people

Linkage to care was defined as visiting a clinic for community HTC and returning to the clinic to obtain CD4 count results (or enrolling in pre-ART care) for facility HTC. Linkage was reported for ten home[14,15,29,34,41–43,65,72,132], six mobile[72,91,92,94,133–135], two campaign[76,77], eight facility VCT[56,81,84,91,92,123,126,136] and five facility PITC studies[60,84,87,111,126]. Home and campaign interventions achieved a high proportion of individuals linked (95%, 95% CI = 87–98%) when paired with facilitated linkage to care strategies (for example, lay-counsellor follow-up to encourage clinic visit); interventions without facilitated linkage achieved lower proportions of HIV-positive individuals visiting a clinic (26%, 95% CI = 18–36%) (Fig. 4). Mobile HTC achieved linkage rates of 37% (95% CI = 24–51%); rates were highest in two interventions conducted in South Africa, one or which used incentivized monetary recruitment and another which used a call centre to encourage linkage after HTC[94,134]. Linkage to care from facility VCT was 61% (95% CI = 48–72%) and from PITC was 55% (95% CI = 39–71%) (Fig. 4). Time from HTC to linkage to care ascertainment varied by study (ranging from 1 to 12 months); the method of ascertainment (participant self-report or clinic record) also varied.

Four home HTC studies reported ART initiation among those eligible[14,41,43,65]. Similar to linkage to care, ART initiation was higher in home interventions with facilitated linkage (76%, 95% CI = 68–82%) compared with those without facilitated linkage (16%, 95% CI = 12–20%) (Fig. 5). ART initiation rates after home HTC with facilitated linkage were similar to those achieved through facility HTC. Initiation among those eligible was 64% (95% CI = 54–72%) in facility VCT and 70% (95% CI = 61–78%) in facility PITC, with 3 studies reporting initiation rates for VCT[61,126,130] and 4 for facility PITC[60,81,84,87,111]. Self-testing showed an ART initiation rate of 29% (95% CI = 17–45%), although this number is among all HIV-positive individuals and is not restricted to those who are ART eligible because point of care CD4 testing was not conducted[11] (Supplementary Figure S55).

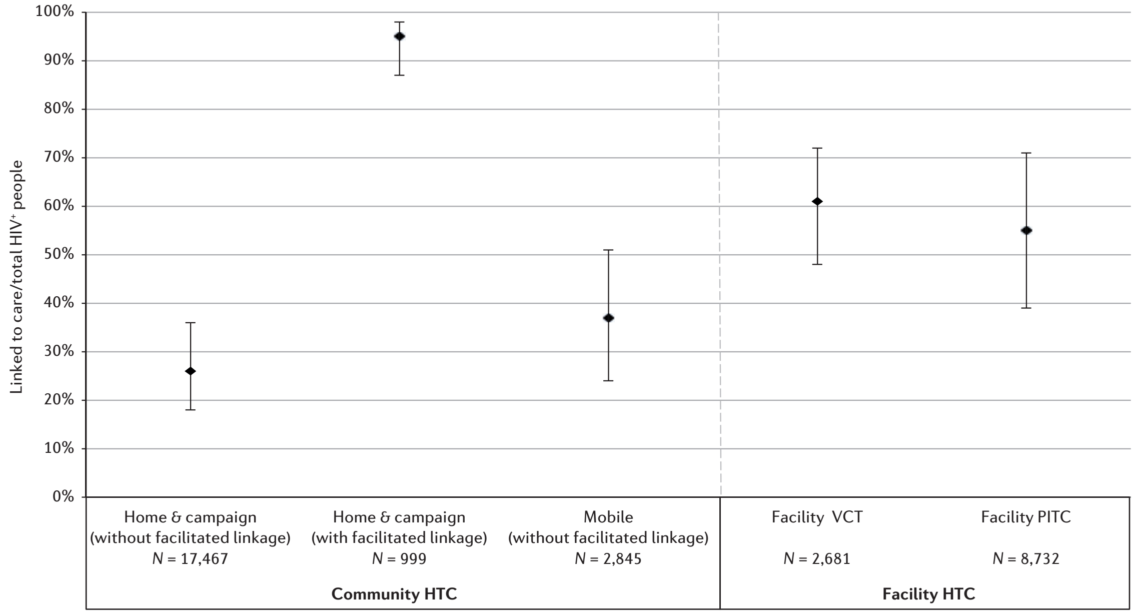


Figure 4 | Linkage to care after community and facility HIV testing and counselling (HTC). Bars indicate 95% confidence intervals of random effects meta-analyses. *N*, sample size. PITC, provider-initiated testing and counselling; VCT, voluntary counselling and testing.

One study reported retention in care at 12 months after ART initiation for home HTC[14] and two studies of both facility VCT and PITC reported retention — one at 6 months[60] and one at 12 months[130]. Not surprisingly, linkage rates were higher in the 6-month compared with the 12-month retention study (Supplementary Figure S59). Retention was highest for home HTC, although the sample size was small (93%, 95% CI = 83–97%) (Fig. 5). Facility VCT achieved 53% (95% CI = 32–71%) retention, and PITC retention achieved 64% (95% CI = 32–90%).

Cost per person tested

The average cost per person tested (2012 US dollars) for community HTC was \$27.38 for mobile, \$16.60 for index, \$11.17 for campaign and \$8.58 for home HTC[88,93,103,137–141] (Supplementary Table S2 and Figure S61). The cost per person tested was highest for stand-alone VCT (\$36.78)[88,93,142]. Hospital and clinic HTC had similar costs (\$12.56 and \$12.32, respectively)[81,88,93,140,142–147] (Supplementary Table S3 and Figure S62). Costs were dependent on the country where the study was conducted, the costs that were included (start-up or ongoing only) and the intervention scale.

DISCUSSION

Across modalities, community HTC successfully reached target groups (men, young adults and first-time testers) with higher coverage than facility HTC (Table 1). High uptake of community HTC reflects acceptability of testing outside of health-care facilities. Community HTC identified HIV-positive individuals with higher CD4 counts who were likely to be earlier in their disease course. Combined with the potential of community HTC with facilitated linkage to achieve high linkage to treatment with similar retention rates as facility HTC, this suggests that scaling up community interventions could reduce the morbidity, mortality and transmission associated with late or non-initiation of ART. Although community interventions test a large number of HIV-negative individuals, HTC can reduce risky sexual behaviour[74] and provide a means to link uninfected persons to primary prevention. This is particularly crucial for young women, who have high HIV incidence and can benefit from PrEP[21]. Preventing HIV infections averts future treatment costs as well as morbidity. A recent modelling study found that ART scale up should be combined with primary prevention such as PrEP to achieve maximum HIV reduction[148]. High coverage of HTC can also reduce stigma around testing.

Each HTC modality reaches distinct subpopulations and a combination of strategies will probably be necessary to achieve high ART coverage. Mobile and campaign HTC had high uptake (97%), as individuals who present at a mobile van or during a campaign are probably seeking out testing, but home HTC also achieved high uptake among people who were offered testing (82%). Home HTC also attained high population coverage, probably because offering testing door-to-door removes substantial barriers, including eliminating the need to actively seek out HIV testing[149]. However, home HTC is less likely to reach men and young adults. A recent home HTC intervention in Botswana reached 85% of women in the target population compared with just 50% of men[150]. This may be because women are more likely to be home at times when the intervention is conducted.

Campaign HTC has the potential to attain high coverage in large catchment areas and identify HIV-positive individuals at high CD4 counts (one-third of newly diagnosed HIV-positive individuals had a CD4 count of 350 cells μl^{-1} or less compared with two-thirds or more for facility HTC). The multidisease focus of campaigns may reduce stigma of HIV testing interventions. Our results suggest that campaign HTC can be a successful strategy for countries seeking to increase overall testing coverage in a short time frame.

Home HTC with self-testing reached the greatest proportion of young adults of all modalities examined[11] and is a promising strategy with high uptake[151]. Young adults (age 15 to 24 years) represent 39% of new infections in those over 15 years old[23], but have lower access to HTC and HIV care and poorer clinical outcomes than other age groups[24]. Home HTC with self-testing had slightly lower coverage and reached fewer first-time testers than home HTC administered by counsellors. The World Health Organization recommends HIV self-testing as an option for individuals who are unable or unwilling to receive counsellor-administered HTC. However, supervision improves interpretation of results[151] and a reactive self-test should not be considered a definitive diagnosis, as standard testing is needed to confirm results. More studies evaluating linkage to care following a positive self-test are needed[16].

Mobile HTC is the most effective strategy for reaching men — a target group in sub-Saharan Africa. Men are more likely to be lost at each step of the HIV treatment cascade; they are less likely to undergo testing, more likely to start ART at an advanced disease stage and more likely to interrupt treatment — all of which leads to increased morbidity and mortality[22]. Qualitative studies highlight men's preference to test outside of facilities[152], so scale up of community interventions can meet this need. Future studies could investigate HTC at predominantly male workplaces, nightclubs or bars.

Index testing of sexual partners through active contact tracing is an efficient high-yield method that should be scaled up. HIV positivity was 55% in this group and the intervention attained a high coverage (41%). The HIV prevalence we

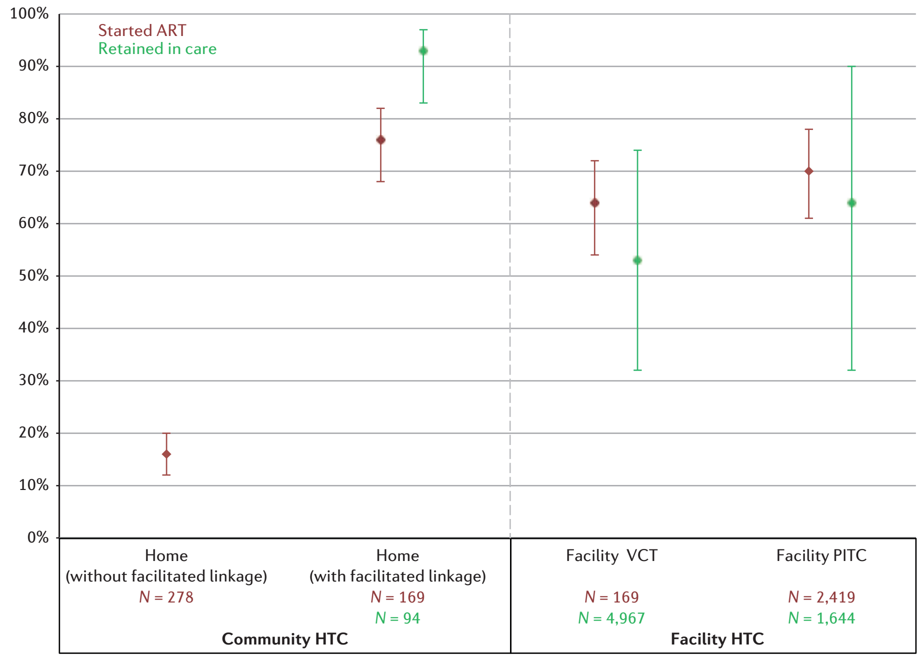


Figure 5 | Pooled percentage initiated antiretroviral therapy (ART) between those eligible and retained in care among those who initiated ART. Bars indicate 95% confidence intervals of random effects meta-analyses. *N*, sample size. PITC, provider-initiated testing and counselling; VCT, voluntary counselling and testing.

report is similar to that found in the literature — 45–50% in cohabitating partners of HIV-positive adults, most of whom are unaware of their status[48]. Interestingly, high coverage of males was achieved only through active contact tracing, whereas passive tracing identified more women (Supplementary Figure S18).

Facilitated linkage strategies are a key component of successful community-based HTC. Individuals testing at an HIV facility generally have higher rates of linking to care and initiating ART than those who test outside the health-care system. However, we found that high linkage rates (comparable with, or higher than, facility HTC) can be achieved with community HTC when individuals are followed-up to encourage linkage.

Although scaling up community HTC with facilitated linkage is important, the benefits of improving facility HTC coverage should not be overlooked. Consistent with previous studies, our analysis finds opt-out facility PITC had much greater uptake than referring patients to VCT[56]. However, coverage of PITC in health facilities is low, demonstrating missed opportunities to identify HIV-positive individuals and to link them to care. For example, a Ugandan hospital reported only 50% of inpatients with HIV-related diagnoses were tested for HIV before leaving the hospital[86]. PITC is an underused strategy in sub-Saharan Africa and scaling up testing would provide a safety net for those who do not independently seek HTC[61,112]. Because PITC identifies mainly symptomatic HIV-positive individuals with low CD4 counts as well as those with health-care access, it should be coupled with other modalities to maximize population coverage.

Our review identified gaps where additional evidence is needed. A large proportion of CD4-count and linkage data came from South Africa, with Uganda and Kenya also well represented. South Africa has the lowest percentage of first-time testers, reflecting the successful scale-up of HTC. There are fewer studies from other parts of sub-Saharan Africa, which may limit how much the pooled estimates can be generalized. Also, few studies followed patients longitudinally and measured linkage to care, ART initiation, retention and viral suppression. In addition, although many studies evaluated home HTC, more data are needed for other community modalities, including campaign and workplace.

Data were also limited for key populations. Despite having an HIV prevalence up to eight times higher than the general population, interventions for key populations are scarce and scale up is urgently needed[115,153]. Key population interventions can reduce the spread of HIV in the general population[154]. Currently, numerous policy barriers exist that restrict the availability and access of HIV-related services for MSM and CSWs, including police harassment and criminal laws[155]. Only three HTC interventions were targeted to MSM and only one was targeted to CSWs and PWIDs. Most key population HTC studies were from Nigeria; data are needed from other parts of sub-Saharan Africa. We report a high HIV positivity combined with a high proportion of first-time testers in MSM and CSW groups, highlighting the need for service expansion. We found a lower HIV prevalence in PWIDs compared with MSM and CSW groups, reflecting sexual transmission as the main mode of HIV spread in sub-Saharan Africa. Successful HTC programmes for key populations are community based (particularly mobile) as many high-risk groups are marginalized and do not have access to conventional health systems[122]. Community-based HTC for MSM and PWIDs have been shown to have higher acceptance and greater HIV yield than clinic referral for HTC[115]. In addition, self-testing is a potential strategy to reach key populations, as it demonstrates high acceptability and is considered convenient and private[156].

Costs of community-based and facility-based HTC vary by modality, country, scale of intervention, linkage strategy and costs included. Generally, community-based HTC and integrated facility HTC costs were comparable. However, stand-alone HTC had the highest cost per person tested, indicating that integrated HTC may be more cost-efficient than stand-alone services (Supplementary Table S3).

The limitations of our analysis included the heterogeneity across studies, which may not be accurately reflected in the pooled estimates. Differences in study design, geographical location (country, urban or rural area) and intervention year added to the heterogeneity. To address this, we used random effects meta-analysis and stratified on key variables (year <2005, country and facilitated linkage). In addition, large numbers of HIV-positive individuals were lost to follow-up in studies that reported linkage, so we considered these individuals unlinked in our analyses. If individuals linked at another clinic, our estimates may be conservative[157]. Furthermore, assessment of linkage to care differed by study (self-report or clinic records review), as did time to linkage assessment, which varied from 1 to 12 months after HTC. In addition, CD4 count at diagnosis and ART uptake among those with eligible CD4 counts could only be assessed in community HTC interventions employing point-of-care CD4, as studies that report CD4 only for those visiting a clinic would not provide accurate denominators. Only studies reporting linkage to care among those eligible for ART were included in our main analysis. Also, estimates of coverage vary in their precision because some studies conducted population enumeration and others used census estimates of the catchment area. Finally, proportion of first-time testers, men and young adults tested are crude measures of relative uptake. For example, for home HTC, it is not possible to discern whether the 40% of those tested being

Table 1 | Summary of HIV testing and counselling coverage and tester demographics.

Parameter	Home		Mobile		Self-testing (home)		Campaign		Index		Key populations		Facility VCT		Facility PITC	
	%	95% CI	%	95% CI	%	95% CI	%	95% CI	%	95% CI	%	95% CI	%	95% CI	%	95% CI
Coverage (accepted/ target population)	70	58–79	37	33–42			76	49–95	41	26–57			15	9–21	18	8–31
Uptake (accepted/ offered)	82	76–87	97	94–99	69	59–78	97	93–99	50	31–71			26	15–39	73	55–87
Young adult (age <25 or 30)	49	43–54	51	44–58	66	65–67	31	12–54					46	39–53	38	24–54
Men	40	39–42	50	47–54	44	42–48	41	37–46	41	23–61			41	38–44	42	39–46
First-time testers	58	48–67	63	50–74			55	28–81			83	71–91	53	40–66	55	48–62
CD4 ≤350 cells μl^{-1}	39	32–46	38	36–41			26	22–30					66	60–72	71	67–75
HIV positivity	10	8–12	11	8–13	8	5–11	6	4–10	55	49–61	16	9–26	18	13–23	20	17–24

CI, confidence interval; PITC, provider-initiated testing counselling; VCT, voluntary counselling and testing

men reflects a lower coverage of men, or a greater coverage of women, or a combination of the two. Future studies reporting the number of men, first-time testers and young adults offered testing compared with those accepting testing would increase the accuracy of these measures. Our findings on uptake, HIV positivity and CD4 count at diagnosis are similar to a previously published meta-analysis[9].

This analysis characterizes linkage and populations reached by HTC modalities to inform policymakers who are charged with addressing gaps in testing. Facility HTC, although important, is unlikely to be sufficient to curb the HIV epidemic because many people in sub-Saharan Africa do not have regular access to health care. Scaling a combination of community HTC, mobile testing to reach men, self-testing to reach young adults and outreach to high-risk populations, as appropriate to the local epidemic setting, is crucial to achieve high knowledge of serostatus and linkage to HIV treatment and prevention in sub-Saharan Africa.

1. Iwuji, C. C. *et al.* Evaluation of the impact of immediate versus WHO recommendations-guided antiretroviral therapy initiation on HIV incidence: the ANRS 12249 TasP (Treatment as Prevention) trial in Hlabisa sub-district, KwaZulu-Natal, South Africa: study protocol for a cluster randomised controlled trial. *Trials* **14**, 230 (2013).
2. Kranzer, K., Govindasamy, D., Ford, N., Johnston, V. & Lawn S. D. Quantifying and addressing losses along the continuum of care for people living with HIV infection in sub-Saharan Africa: a systematic review. *J. Inter. AIDS Soc.* **15**, 17383 (2012).
3. UNICEF. *UGANDA Fast Facts* http://www.unicef.org/uganda/UNICEF_UGANDA_FAST_FACTS_July_2012.pdf (UNICEF, 2012).
4. UNAIDS. *Fast-Track: Ending the AIDS Epidemic by 2030* http://www.unaids.org/sites/default/files/media_asset/JC2686_WAD2014report_en.pdf (UNAIDS, 2014).
5. Fylkesnes, K., Sandoy, I. F., Jurgensen, M., Chipimo, P.J., Mwangala, S. & Michelo, C. Strong effects of home-based voluntary HIV counselling and testing on acceptance and equity: a cluster randomised trial in Zambia. *Social Sci. Med.* **86,** 9–16 (2013).
6. Musheke, M. *et al.* A systematic review of qualitative findings on factors enabling and deterring uptake of HIV testing in Sub-Saharan Africa. *BMC Public Health* **13,** 220 (2013).
7. Siedner, M. J, Ng, C. K., Bassett, I. V., Katz, I. T., Bangsberg, D. R. & Tsai, A. C. Trends in CD4 count at presentation to care and treatment initiation in sub-Saharan Africa, 2002–2013: a meta-analysis. *Clin. Infect. Dis.* **60,** 1120–1127 (2015).
8. Paulin, H. N. *et al.* HIV testing service awareness and service uptake among female heads of household in rural Mozambique: results from a province-wide survey. *BMC Public Health* **15,** 1388 (2015).
9. Suthar, A. B. *et al.* Towards universal voluntary HIV testing and counselling: a systematic review and meta-analysis of community-based approaches. *PLoS Med.* **10,** e1001496 (2013).
10. van Rooyen, H. *et al.* Mobile VCT: reaching men and young people in urban and rural South African pilot studies (NIMH Project Accept, HPTN 043). *AIDS Behav.* **17,** 2946–2953 (2013).
11. MacPherson, P. *et al.* Effect of optional home initiation of HIV care following HIV self-testing on antiretroviral therapy initiation among adults in Malawi: a randomized clinical trial. *J. Am. Med. Assoc.* **312,** 372–379 (2014).
12. Ahmed, S. *et al.* HIV counseling and testing and access-to-care needs of populations most-at-risk for HIV in Nigeria. *AIDS Care* **25,** 85–94 (2013).
13. Coates, T. J. *et al.* Effect of community-based voluntary counselling and testing on HIV incidence and social and behavioural outcomes (NIMH Project Accept; HPTN 043): a cluster-randomised trial. *Lancet Glob. Health* **2,** e267–e277 (2014).
14. Barnabas, R. V. *et al.* Initiation of antiretroviral therapy and viral suppression after home HIV testing and counselling in KwaZulu-Natal, South Africa, and Mbarara district, Uganda: a prospective, observational intervention study. *Lancet HIV* **1,** e68–e76 (2014).
15. Genberg, B. L. *et al.* Linkage to and engagement in HIV care in western Kenya: an observational study using population-based estimates from home-based counselling and testing. *Lancet HIV* **2,** e20–e26 (2015).
16. World Health Organization. *Consolidated Guidelines on HIV Testing Services* http://www.who.int/hiv/pub/guidelines/hiv-testing-services/en/ (WHO, 2015).
17. UNAIDS. *The Gap Report* http://www.unaids.org/sites/default/files/media_asset/UNAIDS_Gap_report_en.pdf (UNAIDS, 2015).
18. Sekandi, J. N. *et al.* High acceptance of home-based HIV counseling and testing in an urban community setting in Uganda. *BMC Public Health* **11,** 730 (2011).
19. Rosenberg, N. E. *et al.* Assessing the effect of HIV counselling and testing on HIV acquisition among South African youth. *AIDS* **27,** 2765–2773 (2013).
20. Hoffmann, C. J. *et al.* Changing predictors of mortality over time from cART start: implications for care. *J. AIDS* **58,** 269–276 (2011).
21. Mirkuzie, A. H., Sisay, M. M., Moland, K. M. & Astrom, A. N. Applying the theory of planned behaviour to explain HIV testing in antenatal settings in Addis Ababa — a cohort study. *BMC Health Services Res.* **11,** 196 (2011).
22. Cornell, M., McIntyre, J. & Myer, L. Men and antiretroviral therapy in Africa: our blind spot. *Trop. Med.Internat. Health* **16,** 828–829 (2011).
23. Kurth, A. E., Lally, M. A., Choko, A.T., Inwani, I. W. & Fortenberry, J. D. HIV testing and linkage to services for youth. *J. Int. AIDS Soc.* **18,** 19433 (2015).
24. MacPherson, P. *et al.* Service delivery interventions to improve adolescents' linkage, retention and adherence to antiretroviral therapy and HIV care. *Trop. Med. Int. Health* **20,** 1015–1032 (2015).
25. Mine, M. *et al.*. Performance of rapid HIV testing by lay counselors in the field during the behavioral and biological surveillance survey among female sex workers and men who have sex with men in Botswana. *J. Acquir. Immune Defic. Syndr.* **68,** 365–368 (2015).
26. Liberati, A. *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *Br. Med. J.* **339,** b2700 (2009).
27. Uwimana, J., Zarowsky, C., Hausler, H. & Jackson D. Training community care workers to provide comprehensive TB/HIV/PMTCT integrated care in KwaZulu-Natal: lessons learnt. *Trop. Med. Int. Health* **17,** 488–496 (2012).
28. Rhodes, S. D., Malow, R. M. & Jolly, C. Community-based participatory research: a new and not-so-new approach to HIV/AIDS prevention, care, and treatment. *AIDS Edu. Prev.* **22,** 173–183 (2010).
29. Bigogo, G. *et al.* The impact of home-based HIV counseling and testing on care-seeking and incidence of common infectious disease syndromes in rural western Kenya. *BMC Infect. Dis.* **14,** 376 (2014).
30. Gonzalez, R. *et al.* High HIV prevalence in a southern semi-rural area of Mozambique: a community-based survey. *HIV Med.* **13,** 581–588 (2012).
31. Helleringer, S., Kohler H. P., Frimpong, J. A. & Mkandawire, J. Increasing uptake of HIV testing and counseling among the poorest in sub-Saharan countries through home-based service provision. *J. Acquir. Immune Defic. Syndr.* **51,** 185–193 (2009).
32. Kimaiyo, S. *et al.* Home-based HIV counselling and testing in western Kenya. *East African Med. J.* **87,** 100–108 (2010).
33. Kranzer, K. *et al.* Individual, household and community factors associated with HIV test refusal in rural Malawi. *Trop. Med. Inter. Health* **13,** 1341–1350 (2008).
34. Medley, A. *et al.* Early uptake of HIV clinical care after testing HIV-positive during home-based testing and counseling in western Kenya. *AIDS Behav.* **17,** 224–234 (2013).
35. Michelo, C., Sandoy, I.F., Dzekedzeke, K., Siziya, S. & Fylkesnes, K. Steep HIV prevalence declines among young people in selected Zambian communities: population-based observations (1995-2003). *BMC Public Health* **6,** 279 (2006).
36. Molesworth, A. M. *et al.* High accuracy of home-based community rapid HIV testing in rural Malawi. *J. Acquir. Immune Defic. Syndr.* **55,** 625–630 (2010).
37. Negin, J., Wariero, J., Mutuo, P., Jan, S. & Pronyk, P. Feasibility, acceptability and cost of home-based HIV testing in rural Kenya. *Trop. Med. Inter. Health* **14,** 849–855 (2009).
38. Ng'ang'a, A. *et al.* The status of HIV testing and counseling in Kenya: results from a nationally representative population-based survey. *J. Acquir. Immune Defic. Syndr.* **66** (Suppl 1), S27–S36 (2014).
39. Bailey, R. C., *et al.* Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. *Lancet* **369,** 643–656 (2007).
40. Shisana, O. South African national household survey of HIV/AIDS prevalence, behavioural risks and mass media impact — detailed methodology and response rate results. *S. Afr. Med. J.* **94,** 283–288 (2004).

41. Tumwebaze, H. *et al.* Household-based HIV counseling and testing as a platform for referral to HIV care and medical male circumcision in Uganda: a pilot evaluation. *PloS ONE* **7,** e51620 (2012).
42. Tumwesigye, E., Wana, G., Kasasa, S., Muganzi, E. & Nuwaha, F. High uptake of home-based, district-wide, HIV counseling and testing in Uganda. *AIDS Patient Care STDs* **24,** 735–741 (2010).
43. van Rooyen, H. *et al.* High HIV testing uptake and linkage to care in a novel program of home-based HIV counseling and testing with facilitated referral in KwaZulu-Natal, South Africa. *J. Acquir. Immune Defic. Syndr.* **64,** e1–e8 (2013).
44. Wachira, J., Ndege, S., Koech, J., Vreeman, R. C., Ayuo, P. & Braitstein, P. HIV testing uptake and prevalence among adolescents and adults in a large home-based HIV testing program in Western Kenya. *J. Acquir. Immune Defic. Syndr.* **65,** e58–e66 (2014).
45. Welz, T. *et al.* Continued very high prevalence of HIV infection in rural KwaZulu-Natal, South Africa: a population- based longitudinal study. *AIDS* **21,** 1467–1472 (2007).
46. Chamie, G. *et al.* Uptake of community-based HIV testing during a multi-disease health campaign in rural Uganda. *PloS ONE* **9,** e84317 (2014).
47. Lugada, E. *et al.* Rapid implementation of an integrated large-scale HIV counseling and testing, malaria, and diarrhea prevention campaign in rural Kenya. *PloS ONE* **5,** e12435 (2010).
48. Lugada, E. *et al.* Comparison of home and clinic-based HIV testing among household members of persons taking antiretroviral therapy in Uganda: results from a randomized trial. *J. Acquir. Immune Defic. Syndr.* **55,** 245–252 (2010).
49. Brown, L. B. *et al.* HIV partner notification is effective and feasible in sub-Saharan Africa: opportunities for HIV treatment and prevention. *J. Acquir. Immune Defic. Syndr.* **56,** 437–442 (2011).
50. Armbruster, B., Helleringer, S., Kalilani-Phiri, L., Mkandawire, J. & Kohler, H. P. Exploring the relative costs of contact tracing for increasing HIV case finding in sub-Saharan countries. *J. Acquir. Immune Defic. Syndr.* **58,** e29–36 (2011).
51. Bwambale, F. M., Ssali, S. N., Byaruhanga, S., Kalyango, J. N. & Karamagi, C. A. Voluntary HIV counselling and testing among men in rural western Uganda: implications for HIV prevention. *BMC Public Health* **8,** 263 (2008).
52. Cawley, C. *et al.* Low rates of repeat HIV testing despite increased availability of antiretroviral therapy in rural Tanzania: findings from 2003–2010. *PloS ONE* **8,** e62212 (2013).
53. Fylkesnes, K. & Siziya, S. A randomized trial on acceptability of voluntary HIV counselling and testing. *Trop. Med. Int. Health* **9,** 566–572 (2004).
54. Isingo, R. *et al.* Trends in the uptake of voluntary counselling and testing for HIV in rural Tanzania in the context of the scale up of antiretroviral therapy. *Trop. Med. Int. Health* **17,** e15–25 (2012).
55. MacPherson, P. *et al.* Suboptimal patterns of provider initiated HIV testing and counselling, antiretroviral therapy eligibility assessment and referral in primary health clinic attendees in Blantyre, Malawi. *Trop. Med. Int. Health* **17,** 507–517 (2012).
56. Dalal, S. *et al.* Provider-initiated HIV testing and counseling: increased uptake in two public community health centers in South Africa and implications for scale-up. *PloS ONE* **6,** e27293 (2011).
57. Fetene, N. W. & Feleke, A. D. Missed opportunities for earlier HIV testing and diagnosis at the health facilities of Dessie town, North East Ethiopia. *BMC Public Health* **10,** 362 (2010).
58. Kayigamba, F. R. *et al.* Provider-initiated HIV testing and counselling in Rwanda: acceptability among clinic attendees and workers, reasons for testing and predictors of testing. *PloS ONE* **9,** e95459 (2014).
59. Kharsany, A. B., Karim, Q.A. & Karim, S. S. Uptake of provider-initiated HIV testing and counseling among women attending an urban sexually transmitted disease clinic in South Africa — missed opportunities for early diagnosis of HIV infection. *AIDS Care* **22,** 533–537 (2010).
60. Topp, S. M. *et al.* Opt-out provider-initiated HIV testing and counselling in primary care outpatient clinics in Zambia. *Bull. WHO* **89,** 328-335 (2011).
61. Topp, S. M. *et al.* Does provider-initiated counselling and testing (PITC) strengthen early diagnosis and treatment initiation? Results from an analysis of an urban cohort of HIV-positive patients in Lusaka, Zambia. *J. Int. AIDS Soc.* **15,** 17352 (2012).
62. Angotti, N. *et al.* Increasing the acceptability of HIV counseling and testing with three C's: convenience, confidentiality and credibility. *Social Sci. Med.* **68,** 2263–2270 (2009).
63. Cherutich, P. *et al.* Lack of knowledge of HIV status a major barrier to HIV prevention, care and treatment efforts in Kenya: results from a nationally representative study. *PloS ONE* **7,** e36797 (2012).
64. Chirawu, P. *et al.* Acceptability and challenges of implementing voluntary counselling and testing (VCT) in rural Zimbabwe: evidence from the Regai Dzive Shiri Project. *AIDS Care* **22,** 81–88 (2010).
65. Dalal, W. Home-based HIV testing and counseling in rural and urban Kenyan communities. *J. Acquir. Immune Defic. Syndr.* **62,** e47–54 (2013).
66. Doherty, T. *et al.* Effect of home based HIV counselling and testing intervention in rural South Africa: cluster randomised trial. *Br. Med. J.* **346,** f3481 (2013).
67. Hensen, B. *et al.* Factors associated with HIV-testing and acceptance of an offer of home-based testing by men in rural Zambia. *AIDS Behav.* **19,** 492–504 (2015).
68. Maheswaran, H., Thulare, H., Stanistreet, D., Tanser, F. & Newell, M. L. Starting a home and mobile HIV testing service in a rural area of South Africa. *J. Acquir. Immune Defic. Syndr.* **59,** e43–46 (2012).
69. Mutale, W., Michelo, C., Jurgensen, M. & Fylkesnes, K. Home-based voluntary HIV counselling and testing found highly acceptable and to reduce inequalities. *BMC Public Health* **10,** 347 (2010).
70. Naik, R., Tabana, H., Doherty, T., Zembe, W. & Jackson, D. Client characteristics and acceptability of a home-based HIV counselling and testing intervention in rural South Africa. *BMC Public Health* **12,** 824 (2012).
71. Nyigo, V. *et al.* Magnitude of HIV infection among older people in Mufindi and Babati districts of the Tanzania mainland. *HIV/AIDS Res. Palliative Care* **6,** 75–79 (2014).
72. Parker, L. A. *et al.* Feasibility and effectiveness of two community-based HIV testing models in rural Swaziland. *Trop. Med. Int. Health* **20,** 893–902 (2015).
73. Shapiro, A. E. *et al.* Community-based targeted case finding for tuberculosis and HIV in household contacts of patients with tuberculosis in South Africa. *Am. J. Respir. Crit. Care Med.* **185,** 1110–1116 (2012).
74. Wolff, B., Nyanzi, B., Katongole, G., Ssesanga, D., Ruberantwari, A. & Whitworth, J. Evaluation of a home-based voluntary counselling and testing intervention in rural Uganda. *Health Policy Plan.* **20,** 109–116 (2005).
75. Choko, A. T. *et al.* The uptake and accuracy of oral kits for HIV self-testing in high HIV prevalence setting: a cross-sectional feasibility study in Blantyre, Malawi. *PLoS Med.* **8,** e1001102 (2011).
76. Granich, R., Muraguri, N., Doyen, A., Garg, N. & Williams, B. G. Achieving universal access for human immunodeficiency virus and tuberculosis: potential prevention impact of an integrated multi-disease prevention campaign in kenya. *AIDS Res. Treat.* **2012,** 412643 (2012).
77. Labhardt, N. D. *et al.* Home-based versus mobile clinic HIV testing and counseling in rural Lesotho: a cluster-randomized trial. *PLoS Med.* **11,** e1001768 (2014).
78. Corbett, E. L. *et al.* Uptake of workplace HIV counselling and testing: a cluster-randomised trial in Zimbabwe. *PLoS Med.* **3,** e238 (2006).
79. Van der Borght, S. F. *et al.* Long-term voluntary counseling and testing (VCT) uptake dynamics in a multicountry HIV workplace program in sub-Saharan Africa. *AIDS Care* **22,** 195–205 (2010).
80. Bemelmans, M. *et al.* Keeping health staff healthy: evaluation of a workplace initiative to reduce morbidity and mortality from HIV/AIDS in Malawi. *J. Int. AIDS Soc.* **14,** 1 (2011).
81. Bassett, I. V. *et al.* Routine voluntary HIV testing in Durban, South Africa: the experience from an outpatient department. *J. Acquir. Immune Defic. Syndr.* **46,** 181–186 (2007).
82. Abdurahman, S., Seyoum, B. & Oljira, L. Factors affecting acceptance of provider-initiated HIV testing and counseling services among outpatient clients in selected health facilities in harar town, Eastern Ethiopia. *HIV/AIDS Res. Pal. Care* **7,** 157–165 (2015).
83. LaCourse, S. M. *et al.* Implementation of routine counselor-initiated opt-out HIV testing on the adult medical ward at Kamuzu Central Hospital, Lilongwe, Malawi. *J. Acquir. Immune Defic. Syndr.* **69,** e31–e35 (2015).
84. Leon, N., Mathews, C., Lewin, S., Osler, M., Boulle, A. & Lombard, C. A comparison of linkage to HIV care after provider-initiated HIV testing and counselling (PITC) versus voluntary HIV counselling and testing (VCT) for patients with sexually transmitted infections in Cape Town, South Africa. *BMC Health Serv. Res.* **14,** 350 (2014).
85. Moodley, J., Bryan, M., Tunkyi, K. & Khedun, S. M. A clinical audit of provider-initiated hiv counselling and testing in a gynaecological ward of a district hospital in KwaZulu-natal, South Africa. *South Afr. J. Obs. .Gynae.* **20,** 8–11 (2014).
86. Wanyenze, R. K. *et al.* Acceptability of routine HIV counselling and testing, and HIV seroprevalence in Ugandan hospitals. *Bull. World Health Organ.* **86,** 302–309 (2008).
87. Waxman, M. J. *et al.* Initial outcomes of an emergency department rapid HIV testing program in western Kenya. *AIDS Patient Care STDs* **21,** 981–986 (2007).
88. Menzies, N. *et al.* The costs and effectiveness of four HIV counseling and testing strategies in Uganda. *AIDS* **23,** 395–401 (2009).
89. Mulogo, E. M., Abdulaziz, A. S., Guerra, R. & Baine, S. O. Facility and home based HIV counseling and Testing: a comparative analysis of uptake of services by rural communities in southwestern Uganda. *BMC Health Services Res.* **11,** 54 (2011).
90. Wachira, J., Kimaiyo, S., Ndege, S., Mamlin, J. & Braitstein, P. What is the impact of home-based HIV counseling and testing on the clinical status of newly enrolled adults in a large HIV care program in Western Kenya? *Clin. Infect. Dis.* **54,** 275–281 (2012).
91. Bassett, I. V. *et al.* Linkage to care following community-based mobile HIV testing compared with clinic-based testing in Umlazi Township, Durban, South Africa. *HIV Med.* **15,** 367–372 (2014).
92. Bassett, I. V. *et al.* Finding HIV in hard to reach populations: mobile HIV testing and geospatial mapping in Umlazi Township, Durban, South Africa. *AIDS Behav.* **19,** 1888–1895 (2015).
93. Grabbe, K. L. *et al.* Increasing access to HIV counseling and testing through mobile services in Kenya: strategies, utilization, and cost-effectiveness. *J. Acquir. Immune Defic. Syndr.* **54**, 317–323 (2010).
94. Kranzer, K. *et al.* Incentivized recruitment of a population sample to a mobile HIV testing service increases the yield of newly diagnosed cases, including those in need of antiretroviral therapy. *HIV Med.* **13,** 132–137 (2012).
95. Mabuto, T. *et al.* Four models of HIV counseling and testing: utilization and test results in South Africa. *PloS ONE* **9,** e102267 (2014).
96. Meehan, S. A., Naidoo, P., Claassens, M. M., Lombard, C. & Beyers, N. Characteristics of clients who access mobile compared to clinic HIV counselling and testing services: a matched study from Cape Town, South Africa. *BMC Health Serv. Res.* **14,** 658 (2014).
97. Morin, S. F. *et al.* Removing barriers to knowing HIV status: same-day mobile HIV testing in Zimbabwe. *J. Acquir. Immune Defic. Syndr.* **41,** 218–224 (2006).
98. Sweat, M. *et al.* Community-based intervention to increase HIV testing and case detection in people aged 16–32 years in Tanzania, Zimbabwe, and Thailand (NIMH Project Accept, HPTN 043): a randomised study. *Lancet Infect. Dis.* **11,** 525–532 (2011).
99. Van Rie, A. *et al.* High uptake of systematic HIV counseling and testing and TB symptom screening at a primary care clinic in South Africa. *PLoS ONE* **9,** e105428 (2014).
100. Kalibala, S., Tun, W., Cherutich, P., Nganga, A., Oweya, E. & Oluoch, P. Factors associated with acceptability of HIV self-testing among health care workers in Kenya. *AIDS Behav.* **18** (Suppl. 4), S405–414 (2014).
101. Pant Pai, N. *et al.* Will an unsupervised self-testing strategy for HIV work in health care workers of South Africa? A cross sectional pilot feasibility study. *PloS ONE* **8,** e79772 (2013).
102. Arendt, V. *et al.* Clinical screening for HIV in a health centre setting in urban Kenya: an entry point for voluntary counselling, HIV testing and early diagnosis of HIV infection? *Trop. Doc.* **37,** 45–47 (2007).

103. Bassett, I. V. *et al.* Mobile HIV screening in Cape Town, South Africa: clinical impact, cost and cost-effectiveness. *PloS ONE* **9,** e85197 (2014).
104. Creek, T. L. *et al.* Botswana's Tebelopele voluntary HIV counseling and testing network: use and client risk factors for HIV infection, 2000-2004. *J. Acquir. Immune Defic. Syndr.* **43,** 210–218 (2006).
105. Fiscus, S. A. *et al.* Rapid, real-time detection of acute HIV infection in patients in Africa. *J. Infect. Dis.* **195,** 416–424 (2007).
106. Mwangi, M. *et al.* Factors associated with uptake of HIV test results in a nationally representative population-based AIDS indicator survey. *Open AIDS J.* **8,** 7–16 (2014).
107. van Schaik, N., Kranzer, K., Wood, R. & Bekker, L. G. Earlier HIV diagnosis — are mobile services the answer? *S. Afr. Med. J.* **100,** 671–674 (2010).
108. Abdallah, T. M., Ali, A.A. & Adam, I. Provider-initiated HIV testing and counseling among tuberculosis patients in Kassala, Eastern Sudan. *J. Infect. Public Health* **5**, 63–66 (2012).
109. Ansa, G. A., Walley, J. D., Siddiqi, K. & Wei, X. Delivering TB/HIV services in Ghana: a comparative study of service delivery models. *Trans. R. Soc. Trop. Med. Hyg.* **108,** 560–567 (2014).
110. Bondo, M., Modiba, M. C. & Becker, P. HIV infection in general surgical patients at the Ga-Rankuwa/MEDUNSA complex South Africa. *E. Afr. Med. J.* **78,** 395–397 (2001).
111. Kiene, S. M., *et al.* Initial outcomes of provider-initiated routine HIV testing and counseling during outpatient care at a rural Ugandan hospital: risky sexual behavior, partner HIV testing, disclosure, and HIV care seeking. *AIDS Patient Care STDs* **24,** 117–126 (2010).
112. Silvestri, D. M. *et al.* A comparison of HIV detection rates using routine opt-out provider-initiated HIV testing and counseling versus a standard of care approach in a rural African setting. *J. Acquir. Immune Defic. Syndr.* **56,** e9–32 (2011).
113. O'Laughlin, K. N. *et al.* Clinic-based routine voluntary HIV testing in a refugee settlement in Uganda. *J. Acquir. Immune Defic. Syndr.* **67,** 409–413 (2014).
114. Nglazi, M. D., van Schaik, N., Kranzer, K., Lawn, S. D., Wood, R. & Bekker, L. G. An incentivized HIV counseling and testing program targeting hard-to-reach unemployed men in Cape Town, South Africa. *J. Acquir. Immune Defic. Syndr.* **59,** e28–34 (2012).
115. Adebajo, S. *et al.* Evaluating the effect of HIV prevention strategies on uptake of HIV counselling and testing among male most-at-risk-populations in Nigeria; a cross-sectional analysis. *Sex Transm. Infect.* http://dx.doi.org/10.1136/sextrans-2014-051659 (2015)
116. Charurat, M. E. *et al.* Uptake of treatment as prevention for HIV and continuum of care among HIV-positive men who have sex with men in Nigeria. *J. Acquir. Immune Defic. Syndr.* **68** (Suppl 2), S114–123 (2015).
117. Nyblade, L. C. *et al.* Population-based HIV testing and counseling in rural Uganda: participation and risk characteristics. *J. Acquir. Immune Defic. Syndr.* **28,** 463–470 (2001).
118. Arthur, G. R. *et al.* The role for government health centers in provision of same-day voluntary HIV counseling and testing in Kenya. *J. Acquir. Immune Defic. Syndr.* **40,** 329–335 (2005).
119. Gresenguet, G. *et al.* Voluntary HIV counseling and testing: experience among the sexually active population in Bangui, Central African Republic. *J. Acquir. Immune Defic. Syndr.* **31,** 106–114 (2002).
120. Hood, J. E. *et al.* Client characteristics and gender-specific correlates of testing HIV positive: a comparison of stand alone center versus mobile outreach HIV testing and counseling in Botswana. *AIDS Behav.* **16,** 1902–1916 (2012).
121. Corbett, E. L. *et al.* HIV incidence during a cluster-randomized trial of two strategies providing voluntary counselling and testing at the workplace, Zimbabwe. *AIDS* **21,** 483–489 (2007).
122. Mulongo, S. *et al.* Applying innovative approaches for reaching men who have sex with men and female sex workers in the Democratic Republic of Congo. *J. Acquir. Immune Defic. Syndr.* **68** (Suppl. 2)**,** S248–251 (2015).
123. Appiah, L. T., Havers, F., Gibson. J., Kay, M., Sarfo, F. & Chadwick, D. Efficacy and acceptability of rapid, point-of-care HIV testing in two clinical settings in Ghana. *AIDS Patient Care STDs* **23,** 365–369 (2009).
124. Kouassi-M'Bengue, A. *et al.* Co-infection of HIV and HBV in voluntary counseling and testing center in Abidjan. *Asian Pacific J. Trop. Dis.* **1,** 275–278 (2011).
125. Akhigbe, R. E. & Bamidele, J. O. Prevalence and pattern of utilization of voluntary counseling and testing services and HIV infection in Ogbomoso, southwestern Nigeria. *J. Nat. Sci. Biol. Med.* **4,** 163–166 (2013).
126. Kikaya, V. *et al.* Voluntary medical male circumcision programs can address low HIV testing and counseling usage and ART enrollment among young men: lessons from Lesotho. *PloS ONE* **9,** e83614 (2014).
127. Govender, S. *et al.* CD4 counts and viral loads of newly diagnosed HIV-infected individuals: implications for treatment as prevention. *PloS ONE* **9,** e90754 (2014).
128. Tchendjou, P. T. *et al.* Factors associated with history of HIV testing among pregnant women and their partners in Cameroon: baseline data from a Behavioral Intervention Trial (ANRS 12127 Prenahtest). *J. Acquir. Immune Defic. Syndr.* **57** (Suppl. 1)**,** S9–15 (2011).
129. Agaba, P. A. *et al.* Patients who present late to HIV care and associated risk factors in Nigeria. *HIV Med.* **15,** 396–405 (2014).
130. Clouse, K. *et al.* Impact of systematic HIV testing on case finding and retention in care at a primary care clinic in South Africa. *Trop. Med. Int. Health* **19,** 1411–1419 (2014).
131. Haskew, J., Turner, K., Ro, G., Ho, A., Kimanga, D. & Sharif, S. Stage of HIV presentation at initial clinic visit following a community-based HIV testing campaign in rural Kenya. *BMC Public Health* **15,** 16 (2015).
132. Naik, R. *et al.* Linkage to care following a home-based HIV counselling and testing intervention in rural South Africa. *J. Int. AIDS Soc.* **18,** 19843 (2015).
133. Govindasamy, D. *et al.* Linkage to HIV care from a mobile testing unit in South Africa by different CD4 count strata. *J. Acquir. Immune Defic. Syndr.* **58,** 344–352 (2011).
134. van Zyl, M. A., Brown, L. L. & Pahl, K. Using a call center to encourage linkage to care following mobile HIV counseling and testing. *AIDS Care* **27,** 921–925 (2015).
135. Larson, B. A. *et al.* Rapid point-of-care CD4 testing at mobile HIV testing sites to increase linkage to care: an evaluation of a pilot program in South Africa. *J. Acquir. Immune Defic. Syndr.* **61,** e13–17 (2012).
136. Larson, B. A. *et al.* Lost opportunities to complete CD4[+] lymphocyte testing among patients who tested positive for HIV in South Africa. *Bull. World Health Organ.* **88,** 675–680 (2010).
137. Terris-Prestholt, F. *et al.* The role of community acceptance over time for costs of HIV and STI prevention interventions: analysis of the Masaka Intervention Trial, Uganda, 1996–1999. *Sex. Trans. Dis.* **33,** S111–S116 (2006).
138. Kahn, J. G. *et al.* Cost of community integrated prevention campaign for malaria, HIV, and diarrhea in rural Kenya. *BMC Health Services Res.* **11,** 346 (2011).
139. Baral, S. *et al.* Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis. *Lancet Infect. Dis.* **12,** 538–549 (2012).
140. Mulogo, E. M., Batwala, V., Nuwaha, F., Aden, A. S. & Baine, O. S. Cost effectiveness of facility and home based HIV voluntary counseling and testing strategies in rural Uganda. *Afr. Health Sci.* **13,** 423–429 (2013).
141. McConnel, C. E. *et al.* The cost of a rapid-test VCT clinic in South Africa. *S. Afr. Med. J.* **95,** 968–971 (2005).
142. Sweat, M. *et al.* Cost-effectiveness of voluntary HIV-1 counselling and testing in reducing sexual transmission of HIV-1 in Kenya and Tanzania. *Lancet* **356,** 113–121 (2000).
143. Aliyu, H. B. *et al.* What is the cost of providing outpatient HIV counseling and testing and antiretroviral therapy services in selected public health facilities in Nigeria? *J. Acquir. Immune Defic. Syndr.* **61,** 221–225 (2012).
144. Obure, C. D. *et al.* Optimising the cost and delivery of HIV counselling and testing services in Kenya and Swaziland. *Sex. Transm. Infect.* **88,** 498–503 (2012).
145. Forsythe, S., Arthur, G., Ngatia, G., Mutemi, R., Odhiambo, J. & Gilks, C. Assessing the cost and willingness to pay for voluntary HIV counselling and testing in Kenya. *Health Policy Planning* **17,** 187–195 (2002).
146. Hausler, H. P. *et al.* Costs of measures to control tuberculosis/HIV in public primary care facilities in Cape Town, South Africa. *Bull. World Health Organ.* **84,** 528–536 (2006).
147. Thielman, N. M. *et al.* Cost-effectiveness of free HIV voluntary counseling and testing through a community-based AIDS service organization in Northern Tanzania. *Am. J. Public Health* **96,** 114–119 (2006).
148. Stover, J. *et al.* How can we get close to zero? The potential contribution of biomedical prevention and the investment framework towards an effective response to HIV. *PLoS ONE* **9,** e111956 (2014).
149. Sabapathy, K., Van den Bergh, R., Fidler, S., Hayes, R. & Ford, N. Uptake of home-based voluntary HIV testing in sub-Saharan Africa: a systematic review and meta-analysis. *PLoS Med.* **9,** e1001351 (2012).
150. Novitsky, V. *et al.* Estimated age and gender profile of individuals missed by a home-based HIV testing and counselling campaign in a Botswana community. *J. Int. AIDS Soc.* **18,** 19918 (2015).
151. Choko, A. T. *et al.* Uptake, accuracy, safety, and linkage into care over two years of promoting annual self-testing for HIV in Blantyre, Malawi: a community-based prospective Study. *PLoS Med.* **12,** e1001873 (2015).
152. Leblanc, N. M. & Andes, K. L. An exploration of men's knowledge, attitudes, and perceptions of HIV, HIV risk, and willingness to test for HIV in Yendi District, Northern Ghana. *J. Assoc. Nurses AIDS Care* **26,** 281–295 (2015).
153. Wheeler, T., Wolf, R.C., Kapesa, L., Cheng Surdo, A. & Dallabetta, G. Scaling-up HIV responses with key populations in West Africa. *J. Acquir. Immune Defic. Syndr.* **68** (Suppl. 2), S69–73 (2015).
154. Bekker, L. G. *et al.* Combination HIV prevention for female sex workers: what is the evidence? *Lancet* **385,** 72–87 (2015).
155. Duvall, S. *et al.* Assessment of policy and access to HIV prevention, care, and treatment services for men who have sex with men and for sex workers in Burkina Faso and Togo. *J. Acquir. Immune Defic. Syndr.* **68** (Suppl. 2)**,** S189–197 (2015).
156. Figueroa, C., Johnson, C., Verster, A. & Baggaley, R. Attitudes and acceptability on HIV self-testing among key populations: a literature review. *AIDS Behav.* **19,** 1949–1965 (2015).
157. Nsanzimana, S. *et al.* HIV care continuum in Rwanda: a cross-sectional analysis of the national programme. *Lancet HIV* **2,** e208–e215 (2015).

SUPPLEMENTARY MATERIAL

Is linked to the online version of this paper at: www.nature.com/10.1038/nature16044

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contribution of colleagues from the International Clinical Research Center at the University of Washington (UW), the expert assistance from UW librarians, and the authors who responded to our requests for additional information.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests. Financial support for this publication has been provided by the Bill & Melinda Gates Foundation.

ADDITIONAL INFORMATION

ARTICLE OPEN

Comparison of diagnostics for the detection of asymptomatic *Plasmodium falciparum* infections to inform control and elimination strategies

Lindsey Wu*[1], Lotus L. van den Hoogen*[1], Hannah Slater[2], Patrick G. T. Walker[2], Azra C. Ghani[2], Chris J. Drakeley[1] & Lucy C. Okell[2]

The global burden of malaria has been substantially reduced over the past two decades. Future efforts to reduce malaria further will require moving beyond the treatment of clinical infections to targeting malaria transmission more broadly in the community. As such, the accurate identification of asymptomatic human infections, which can sustain a large proportion of transmission, is becoming a vital component of control and elimination programmes. We determined the relationship across common diagnostics used to measure malaria prevalence — polymerase chain reaction (PCR), rapid diagnostic test and microscopy — for the detection of *Plasmodium falciparum* infections in endemic populations based on a pooled analysis of cross-sectional data. We included data from more than 170,000 individuals comparing the detection by rapid diagnostic test and microscopy, and 30,000 for detection by rapid diagnostic test and PCR. The analysis showed that, on average, rapid diagnostic tests detected 41% (95% confidence interval = 26–66%) of PCR-positive infections. Data for the comparison of rapid diagnostic test to PCR detection at high transmission intensity and in adults were sparse. Prevalence measured by rapid diagnostic test and microscopy was comparable, although rapid diagnostic test detected slightly more infections than microscopy. On average, microscopy captured 87% (95% confidence interval = 74–102%) of rapid diagnostic test-positive infections. The extent to which higher rapid diagnostic test detection reflects increased sensitivity, lack of specificity or both, is unclear. Once the contribution of asymptomatic individuals to the infectious reservoir is better defined, future analyses should ideally establish optimal detection limits of new diagnostics for use in control and elimination strategies.

Nature 528, S86–S93 (3 December 2015), DOI: 10.1038/nature16039

Over the past two decades, considerable progress has been made in reducing the global malaria burden. Between 2000 and 2013 alone, malaria-related mortality decreased by 47% worldwide and 54% in Africa. In addition, more than half of malaria endemic countries are on track to meet global targets to reduce malaria incidence by 75% in 2015 (ref. 1). These achievements are largely due to the widespread use of insecticide-treated nets (ITNs) and highly effective antimalarial treatments. The treatment of symptomatic cases in particular has been enabled by notable advances in the development and deployment of more accurate malaria diagnostics[2,3]. However, efforts to reduce the burden of malaria infections further in the future will require moving beyond the treatment of clinical infections to targeting transmission more broadly in the community. As such, the accurate identification of asymptomatic human infections, which can sustain a large proportion of transmission, is becoming a vital component of control and elimination programmes[2,4].

Community chemotherapy (for example, mass screen and treat (MSAT) or mass drug administration (MDA) programmes) in conjunction with ongoing vector control is an approach under consideration for the interruption of transmission. This is achieved through the direct treatment of potentially infectious individuals. In the case of MSAT strategies, delivering drugs specifically on the basis of positive test results may be considered preferable to presumptive treatment because it provides clear benefit to the recipient and limits excess drug use that may drive antimalarial resistance. However, owing to the insufficient sensitivity of existing field diagnostics used to identify asymptomatic infections, studies have shown that MSAT has limited effect in reducing transmission[5,6].

Measuring parasite infection by microscopy has been the gold standard in malaria research for more than a century and remains relatively widespread as a point-of-care diagnostic in clinical and epidemiological settings. More recently, the advent of rapid diagnostic tests (RDTs), which measure the presence of histidine-rich protein 2 (HRP2) for *Plasmodium falciparum* and/or lactate dehydrogenase for other *Plasmodium* species (pLDH), has expanded the range of diagnostic options. Originally developed to inform clinical treatment, RDTs are increasingly important for epidemiological characterization[7] because of their low cost and field applicability. However, most only have reported detection limits in the range of 100 to 200 parasites per microlitre[8,9] in comparison with around 50 parasites per microlitre by expert microscopy[10].

Over the past three decades, the development of nucleic acid amplification tests has improved the detection limit for malaria infection to less than 1 parasite per microlitre by ultrasensitive quantitative polymerase chain reaction (qPCR)[11,12]. Although these detection thresholds are more appropriate for

*These authors contributed equally. [1]Department of Immunology and Infection, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. [2]MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, Norfolk Place, London W2 1PG, UK. Correspondence should be addressed to: L. W. e-mail lindsey.wu@lshtm.ac.uk.

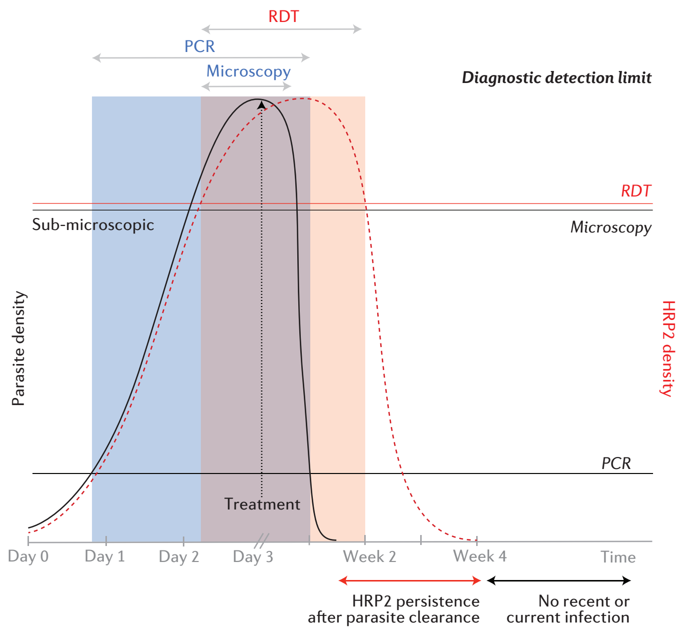


Figure 1 | Schematic of diagnostic detection limits with respect to parasite and HRP2 density. The black curve indicates parasite density and the red curve indicates HRP2 density. Time scale is in days prior to treatment and in weeks after treatment. Horizontal lines are the detection limits of respective diagnostics. The blue shaded area shows detectability of parasites by microscopy and/or polymerase chain reaction (PCR), whereas the red shaded area shows detectability of HRP2 by rapid diagnostic test (RDT).

measuring low-density infections than microscopy and RDTs, most PCR techniques remain impractical for wide-scale use in field surveys owing to cost, processing time and the lack of appropriate laboratory facilities in many endemic countries[10]. Comparative analysis of malaria prevalence, measured by both microscopy and PCR in cross-sectional surveys, has shown that sub-microscopic low-density infections are common across a range of transmission settings[13,14]. These infections may be chronic and asymptomatic, particularly in previously exposed individuals with more mature immune responses. More importantly, even at low parasite densities, they are still capable of infecting mosquitoes and seeding onward transmission[15]. Even though RDTs are becoming more common in areas where these types of infections are prevalent, studies formally evaluating their performance in detecting asymptomatic infections remain scarce.

Recently, there has been an increased focus on developing improved diagnostics to inform malaria elimination strategies. The analysis presented in this paper aims to determine the concordance of current malaria diagnostic methods, forming a baseline to evaluate further how they can be improved to inform malaria control and elimination strategies. It should be noted that, in principle, quantifying the presence of gametocytes is considered the most accurate method for characterizing transmission and the potential infectiousness of individuals. Research in this area is ongoing, but the technical challenges of existing gametocyte assays preclude them from standardized use[16]. Moreover, all malaria infections have the capacity to produce gametocytes[17,18]. Therefore, in the context of community chemotherapy programmes, any individual who tests positive for asexual parasites should be treated to reduce transmission. Given this operational framework, this paper does not address the role of diagnostics that specifically measure gametocytaemia.

So far, no studies have comprehensively evaluated the concordance across PCR, RDT and microscopy detection methods simultaneously in asymptomatic populations. Although microscopy- and PCR-measured prevalences are based on similar biological endpoints (parasite density), diagnostic results based on RDTs are less comparable given that HRP2 and pLDH are indirect measures of parasite biomass[19]. HRP2 can persist in the blood for up to two weeks after parasite clearance[20]. Consequently, results across these diagnostic methods indicate a range of possible infection states, from patent or sub-microscopic infection to recently cleared infection (Fig. 1). A limited number of studies have reviewed the detection capability of RDTs in asymptomatic individuals[8,21], but key research questions still remain. A recent analysis of Demographic and Health Surveys (DHS) across Africa showed a higher prevalence of malaria when measured by RDTs compared with detection by microscopy in 19 out of 22 surveys. This report also highlighted the issue of false positives owing to prolonged presence of HRP2 after parasite clearance[21]. However, studies have not reviewed the detection capability across all three diagnostics. Furthermore, the DHS study only considered children under 5 years of age and did not determine the effect of malaria transmission intensity on diagnostic discordance. This is particularly important given that low-density infections seem to be most common in adults and in low-transmission settings[13,14].

In this study, we determine the relationship across malaria prevalence measures obtained by current diagnostic methods — PCR, RDT and microscopy — for the detection of *P. falciparum* infections in endemic populations based on a pooled analysis of published and unpublished cross-sectional data.

METHODS

Literature review and data collection. We carried out two separate literature reviews to identify studies in which *P. falciparum* prevalence was measured by different diagnostic techniques in the same individuals: first, by RDT and microscopy, and, second, by RDT and PCR. Relevant studies were identified in PubMed and Embase, using MeSH and Map terms when possible. For the RDT and microscopy review, the search terms were: "'rapid diagnostic test' and 'microscopy' [MeSH/Map] and 'malaria falciparum' [MeSH/Map]", and for the RDT and PCR review the search terms were: "'polymerase chain reaction' [MeSH/Map] and 'malaria falciparum' [MeSH/Map]". Searches were limited to English, human and post-2005 (considering the substantial development in RDTs over time[22]). For Embase, the searches were also limited to journal articles. Inclusion criteria were applied as previously described[13]. In short, only studies that were cross-sectional (on populations not selected according to malaria test results or symptoms), that were of populations from a malaria endemic region, that used RDTs targeting *P. falciparum* only or mixed infections (HRP2 and/or pLDH) and that used PCR or loop-mediated isothermal amplification (LAMP) methods were included. For intervention studies, only baseline data were included, except for treatment studies where a sufficient amount of time had passed between last treatment and follow-up. Separate publications that used the same data set or measured 0% prevalence by both methods were removed, as well as data from clusters with fewer than five individuals. RDT and microscopy studies identified in our literature search that also included PCR measurements were included in the RDT and PCR data set, and vice versa for RDT and PCR studies that included microscopy measurements. In addition to the literature review, we sought as many individual-level data sets as possible from studies with the above inclusion criteria.

RDT and microscopy. Where available, information on location, sample size, RDT brand and type (HRP2 or pLDH), age group (15 or younger compared with older than 15) and prevalence estimates were recorded[5,23–42]. Furthermore, data from the DHS online database were extracted[43]. These included individual-level data on location and timing of collection, RDT and microscopy test results, RDT brand[21], age, sex, use of an ITN, fever and antimalarial use in the past two weeks. In addition, individual-level data sets from one unpublished and one published study were included[44], as well as shared data sets of the RDT and PCR comparison that also included microscopy measurements (see below)[45–49].

RDT and PCR. Corresponding authors of the 13 studies identified from the literature search were contacted to request individual-level data in December 2014 and reminders were sent out 4 weeks later. Of the contacted authors, six responded within the timeframe; five data sets were included[45–47,49,50], and one data set had been destroyed for privacy compliance. Prevalence measures and study information (including PCR method) were extracted as described above from the publications in the aforementioned literature search and the non-responders group, as well as included studies from the RDT and microscopy search that also reported PCR proportions[25,27,34,39,40,42,51–55]. Four additional individual-level unpublished and published data sets were included[44,48].

Statistical analyses. We analysed the association between PCR- and RDT-measured prevalence, and microscopy- and RDT-measured prevalence by fitting a linear relationship on the log odds scale[13,56]. Prevalence (on a scale of 0 to 1) was defined as $\frac{e^{(\log odds)}}{1+e^{(\log odds)}}$, where log odds = $\log_e(\frac{prevalence}{1-prevalence})$

$$\Omega_{Ri} = \Omega_{Pi} + \delta_{Ri} \quad (1)$$

$$\delta_{Ri} = \delta'_{Ri} + \beta_0 (\Omega_{Pi} - \bar{\Omega}_P) \quad (2)$$

$$\Omega_{Ri} = \Omega_{Mi} + \delta_{Ri} \quad (3)$$

$$\delta_{Ri} = \delta'_{Ri} + \beta_0 (\Omega_{Mi} - \bar{\Omega}_M) \quad (4)$$

In Equations 1–4, Ω_{Ri} is the log odds of RDT-measured prevalence in trial i, Ω_{Pi} is the log odds of PCR prevalence, Ω_{Mi} is the log odds of microscopy-measured prevalence, δ_{Ri} is the log odds ratio (OR) of RDT- to PCR-measured prevalence (RDT:PCR; Equation 1) or RDT- to microscopy-measured prevalence (RDT:microscopy; Equation 3), δ'_{Ri} is the expected log OR of RDT:PCR prevalence (Equation 2) or RDT:microscopy prevalence (Equation 4) when the log odds of PCR- or microscopy-measured prevalence is equal to the mean across trials, $\bar{\Omega}_P$ and $\bar{\Omega}_M$ are the mean log odds of PCR- and microscopy-measured prevalence, respectively, across trials, and β_0 is the regression coefficient. To allow for varying sample size and sampling variation across the surveys included in our analysis, the model was fitted using Bayesian Markov Chain Monte Carlo methods in JAGS version 3.4.0 and the *rjags* package in R version 3.0.2 (ref. 13). We also explored fitting polynomial relationships, but these provided no substantial improvement in fit to the data over the linear model as assessed by deviance information criterion, nor were these fitted relationships qualitatively different (data not shown). To confirm that the fitted curves at different prevalence ranges were not overly influenced by the high number of data points in lower transmission areas, we fitted separate relationships in three PCR-measured prevalence bands: <5%, 5–20% and >20%. These categories represent approximate cut-offs that have been suggested as thresholds for operational decision-making. Broadly speaking, programmes can begin to consider targeted and focal control strategies when parasite prevalence by microscopy falls below 5% (ref. 57), which translates to a PCR-measured prevalence of 20% (ref. 14), and move towards targeted elimination when it falls below 1% (ref. 58) (5% PCR-measured prevalence[14]).

We also conducted a meta-analysis of the risk ratio between RDT:PCR prevalence or RDT:microscopy prevalence, adjusted for random effects at the study level (for RDT:PCR) or country level (for RDT:microscopy). Studies that reported zero infections by either diagnostic method were assigned a value of 0.01 to allow a risk ratio to be calculated. To evaluate the effect of explanatory factors on discordant test results, individual-level data were analysed by logistic regression, allowing for random effects at the study or country level as noted above. The meta-analysis was done with the *metafor* package in R version 3.0.2, and the logistic regression with the *logit* command in STATA version 13.

We assessed the ability of our models to predict RDT-measured prevalence based on microscopy- or PCR-measured prevalence data. Leave-one-out cross validation was used to evaluate the RDT:PCR and the RDT:microscopy models separately. The data available for direct comparison of malaria detection by RDT and PCR in the same individuals were sparse relative to the quantity of data available for the RDT:microscopy and previous microscopy:PCR comparisons. Therefore, we also triangulated the relationship between RDT- and PCR-measured prevalence by combining the RDT:microscopy relationship calculated in this study with the microscopy:PCR prevalence relationship that has been previously defined[13]. The credible interval of the triangulation line was computed from the posterior distributions of all the parameters from both equations combined. We evaluated whether this triangulated RDT:PCR relationship was significantly different from the observed RDT:PCR relationship using the posterior distributions of the predictions from each model.

RESULTS

Literature search and data collection. The literature search generated 549 results in Pubmed and an additional 37 in Embase for RDT and microscopy, and 2,247 results in PubMed and an additional 426 in Embase for RDT and PCR. In total, 20 RDT: microscopy studies and 13 RDT:PCR studies from the literature search met our inclusion criteria. Combined with additional data sets from DHS and unpublished studies, the pooled data available for evaluation yielded 323 pairs of prevalence estimates for RDT and microscopy[5,23–42,44–49] and 162 pairs for RDT and PCR[25,27,34,39,40,42,45–55]. The extracted proportions together with the main characteristics of the studies from our literature search are provided in the Supplementary Information. The main PCR method used was nested PCR (nPCR; 15 of 20) of which mainly the Snounou method[59] was used (11 of 15). The other methods included LAMP (1 of 20) and qPCR (4 of 20). All of the included RDTs in both comparisons were based on HRP2, with 8 out of 20 studies also including pLDH to measure species other than *P. falciparum*. However, this study only focuses on the detection of *P. falciparum* infections.

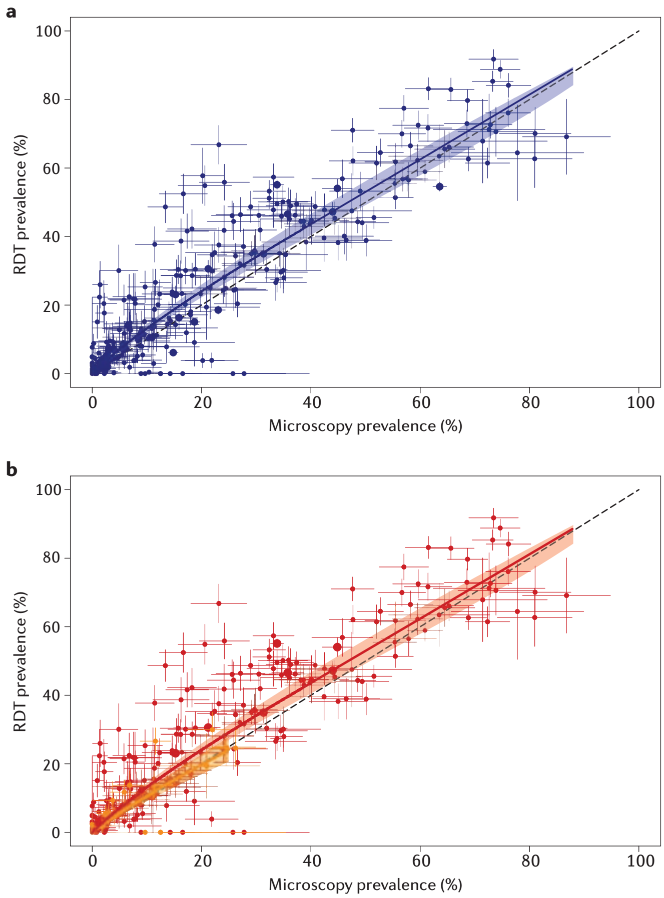


Figure 2 | The relationship between rapid diagnostic tests (RDTs) and microscopy *Plasmodium falciparum* prevalence overall (a) and stratified by age group (b). In **b** red indicates children (those under 15 years) and yellow indicates adults (those over 15 years). Dashed lines indicate the expected relationship if RDT and microscopy detected equal prevalence. Horizontal and vertical lines indicate 95% confidence intervals around point estimates, whereas coloured solid lines indicate the median of the Bayesian posterior distributions from the fitted model and shaded areas indicate 95% credible intervals. Radius of point estimates indicate cluster size (from small to large: <100, 100–1,000 and >1,000).

Comparison of RDT- and microscopy-measured prevalence. Analysis of RDT- and microscopy-measured prevalence included data from 172,281 individuals who were tested with RDTs (cluster prevalence range = 0–92%) and 186,434 tested with microscopy (cluster prevalence range = 0–87%). The 323 geographical clusters spanned a total of 29 countries (cluster size range = 5–7,664). Overall, prevalence of *P. falciparum* measured by microscopy detected 87% (95% confidence interval (CI) = 74–102%) of RDT-positive infections. Therefore, RDT and microscopy detection was comparable (Fig. 2a, Table 1), with less of a difference between the two diagnostic methods in children under 15 years of age (77%, 95% CI = 71–85%) compared with adults (over 15 years)

Table 1 | Best fit relationships between RDT:microscopy and RDT:PCR prevalence.

RDT:microscopy
OVERALL
log odds RDT prevalence = 0.108 + 0.907 × log odds microscopy prevalence (all ages)
BY AGE CATEGORY
log odds RDT prevalence = 0.109 + 0.908 × log odds microscopy prevalence (under 15 years)
log odds RDT prevalence = −0.168 + 0.890 × log odds microscopy prevalence (over 15 years)
RDT:PCR
OVERALL
log odds RDT prevalence = −0.968 + 1.186 × log odds PCR prevalence (all ages)
BY AGE CATEGORY
log odds RDT prevalence = −0.382 + 1.306 × log odds PCR prevalence (under 5 years)
log odds RDT prevalence = −0.864 + 1.213 × log odds PCR prevalence (6–15 years)
log odds RDT prevalence = −1.378 + 1.300 × log odds PCR prevalence (over 15 years)
BY AGE CATEGORY
log odds RDT prevalence = 1.097 + 1.690 × log odds PCR prevalence (<5% prevalence)
log odds RDT prevalence = 0.211 + 1.754 × log odds PCR prevalence (5–20% prevalence)
log odds RDT prevalence = −0.516 + 1.904 × log odds PCR prevalence (>20% prevalence)
PCR PREVALENCE BASED ON DATA TRIANGULATION
log odds PCR prevalence = 0.108 + 0.907 × [(log odds RDT prevalence − 0.954)/0.868]
PCR, polymerase chain reaction; RDT, rapid diagnostic test.

(60%, 95% CI = 48–86%) (Fig. 2b, Table 1). The lower age-specific risk ratios are due to smaller cluster sizes after stratifying the data by age group. However, regression analysis of individual-level data did not show a significant association between age group and test discordance (Supplementary Table 1).

Effect of individual level covariates on RDT:microscopy discordance. In addition to age, we explored the effect of several other covariates on diagnostic outcomes, and adjusted for transmission intensity as assessed by microscopy-measured prevalence (Supplementary Table 1). A significant association was seen between self-reported antimalarial use in the two weeks before survey testing and RDT positivity in individuals who tested negative by microscopy (OR = 1.71, 95% CI = 1.16–2.51, p = 0.006). The presence of fever at the time of testing (recorded temperature with study-specific cut-off or self-reported) reduced the odds of undetected malaria infection by RDT among microscopy-positive individuals (OR = 0.59, 95% CI = 0.39–0.89, p<0.001). Among individuals testing negative by microscopy, presence of a fever was significantly associated with RDT positivity (OR = 1.84, 95% CI = 1.51–2.24, p<0.001), after adjusting for transmission intensity. There was a borderline significant increased risk of malaria infection being undetectable by RDT among those who used an ITN and were microscopy positive (OR = 1.26, 95% CI = 1.00–1.59, p = 0.053), whereas use of an ITN was associated with decreased RDT positivity (OR = 0.84, 95% CI = 0.73–0.97, p = 0.019) among microscopy-negative individuals. There was no evidence of an association between RDT brand and the risk of an undetected malaria infection by RDT among microscopy-positive individuals. Among microscopy-negative individuals, the proportion testing positive was different between RDT brands, but these results are difficult to interpret, owing to complete correlation between study and RDT brand. The year of the survey was not associated with discordant test results for RDT:microscopy.

Comparison of RDT- and PCR-measured prevalence. Analysis of RDT- and PCR-measured prevalence included 35,887 individuals tested with an RDT (cluster prevalence range = 0–45%) and 31,178 individuals tested with PCR (cluster prevalence range = 0–52%). There were a total of 162 geographical clusters across 17 countries (cluster size range = 5–3,307, Figs 3a,b and Table 1). Pooled meta-analysis across all surveys showed that RDTs detected an average of 41% (95% CI = 26–66%) of PCR-positive infections. This primarily reflects the relationship between RDT and PCR in low-transmission settings, with an average PCR prevalence of 8% across all the clusters included in our analysis.

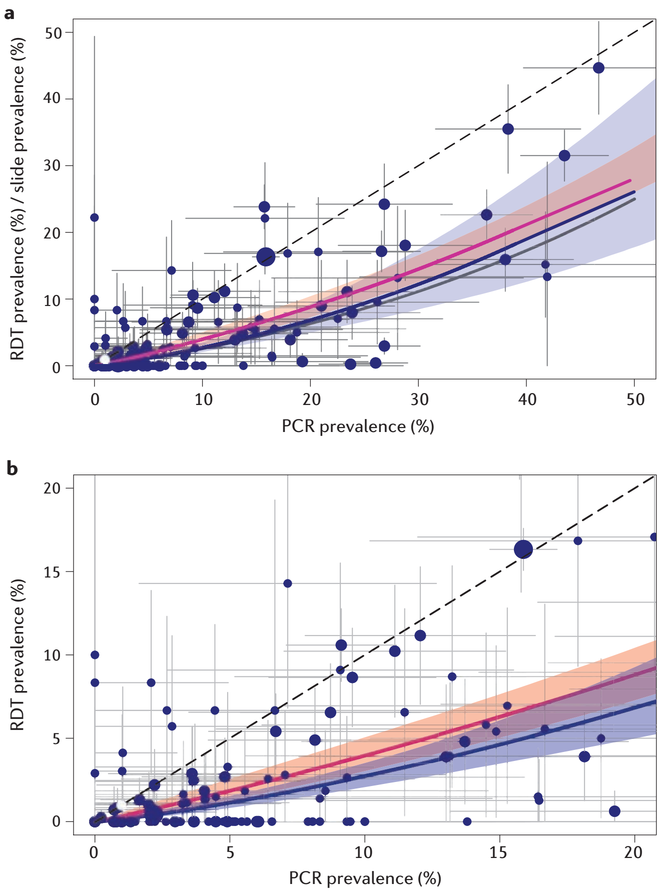


Figure 3 | The relationship between rapid diagnostic test (RDT) and polymerase chain reaction (PCR) prevalence overall (**a**) and zoomed in for <20% PCR prevalence (**b**). Blue, observed RDT:PCR prevalence data and model fit; pink, the triangulated RDT:PCR comparison (see methods); grey, the PCR:microscopy comparison from ref. 13. Dashed lines indicate the expected relationship if RDT (or microscopy) and PCR detected equal prevalence. Horizontal and vertical lines indicate 95% confidence intervals around point estimates, whereas coloured solid lines indicate the median of the Bayesian posterior distributions from the fitted model and shaded areas indicate 95% credible intervals. Radius of point estimates indicate cluster size (from small to large: <100, 100–1,000 and >1,000).

Age, transmission intensity and undetected malaria infection by RDT. As with the relationship between RDT- and microscopy-measured prevalence, stratifying by age group improved the model fit to the data, showing a decrease in detectability by RDT with increasing age (Figs 4a–c). Meta-analysis of the risk ratio between RDT and PCR positivity showed that, for children under 5 years of age, RDTs detected 81% (95% CI = 74–89%) of PCR-positive infections. By comparison, RDTs detected fewer PCR-positive school-aged individuals (6–15 years) (70%, 95% CI = 57–86%), and even fewer among adults over 15 years of age (49%, 95% CI = 31–78%). There was a larger data set available for analysis in the under 5 (140 clusters) and 6–15 (136 clusters) age groups compared with adults (81 clusters), suggesting that additional data in the higher age group could help to improve the accuracy of these estimates.

Previous studies have suggested that the proportion of carriers with sub-microscopic infections decreases in areas of higher transmission intensity, potentially because of an association with re-infection and increased parasite density[13,14]. A similar trend was also observed in the relationship between RDT and PCR detectability. The fit to our data was improved after stratifying by transmission intensity based on PCR-measured prevalence, showing increased RDT sensitivity compared with PCR as transmission increases (Fig. 4d–f). However, meta-analysis of the risk ratio between RDT and PCR positivity did not show a significant difference between the three transmission

ranges, possibly indicating that more data are needed to define a more robust relationship for each transmission setting.

Figure 5 shows RDT detectability as a proportion of PCR-positive individuals, stratified by age and transmission intensity. Irrespective of transmission intensity, adults have the highest percentage of RDT-undetectable infections. By contrast, the percentage of individuals with RDT-detectable infections in all age groups increases as transmission intensity increases. However, since infection rates are greater at high-transmission intensities, RDTs may still miss a larger absolute number of infectious individuals at this level of endemicity. Best-fit model estimates of PCR-measured prevalence based on RDT-measured prevalence are summarized in Figs 3, 4 and Table 1.

Effect of individual-level covariates on RDT:PCR discordance. We evaluated the impact of age and transmission intensity on RDT positivity among PCR-negative individuals as a potential indicator of prolonged HRP2 clearance time. Logistic regression, adjusted for cluster PCR-measured prevalence, showed that among PCR-negative individuals, school-aged children had a significantly higher RDT positivity (OR = 1.53, 95% CI = 1.28–1.82, p<0.001) when compared with a baseline of children under 5 years of age. Adults showed similar odds of being RDT positive (OR = 1.00, 95% CI = 0.64–1.58, p = 0.990) as those under 5 years. Infections that were undetected by RDT, based on PCR positivity, were highest in adults (OR = 5.04, 95 %CI = 4.14–6.13, p<0.001) compared with those under 5 years, with a similar risk in school-aged children and those under 5 years (Supplementary Table 2).

RDT positivity among PCR-negative individuals varied between RDT brands, as did the detection of infection in PCR-positive individuals, but these results were not significant. Patients with a fever were less likely to have undetected infections by RDT if they were PCR positive (OR = 0.14, 95% CI = 0.06–0.32, p<0.001), but also more likely to have a RDT-positive result if they were PCR negative (OR = 4.86, 95% CI = 2.29–10.30, p<0.001). More recent surveys showed a lower risk of RDT-undetected infections, based on PCR positivity (OR = 0.77 per year, 95% CI = 0.60–0.99, p = 0.044), which may indicate an improved performance of RDTs over time. PCR method was associated with test discordance at borderline significance, with RDTs detecting less PCR positive results measured by qPCR than those measured by PCR (OR = 1.92, 95% CI = 0.98–3.74, p = 0.056), reflecting higher sensitivity of qPCR, as described previously[15,45].

Model validation. From the leave-one-out analysis, the correlation coefficient between observed and predicted values of RDT-measured prevalence from the RDT:PCR model was 0.67, indicating a moderate agreement. The correlation coefficient between observed and predicted values of RDT-measured prevalence from the RDT:microscopy model was 0.92, indicating a relatively stronger agreement (Fig. 6). The credible interval of this triangulated relationship was narrower than that of the directly observed line, owing to the larger number of data points in the RDT:microscopy and microscopy:PCR data sets (Figs 2, 3, Table 1). There was no significant difference between the triangulated and observed relationships at any transmission intensity.

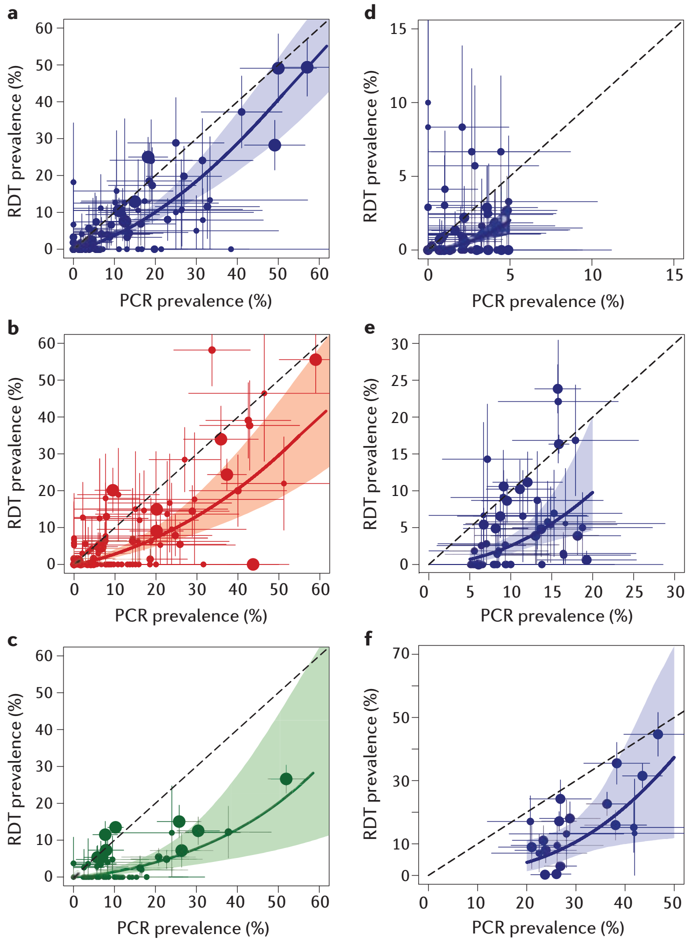


Figure 4 | The relationship between rapid diagnostic test (RDT) and polymerase chain reaction (PCR) prevalence by age group (**a**–**c**) and PCR prevalence band (**d**–**f**). The Bayesian model was fitted separately for each age group or PCR prevalence band. Age groups are younger than 5 years (**a**) 6–15 years (**b**) and older than 15 years (**c**). PCR prevalence bands are <5% (**d**), 5–20% (**e**) and >20% (**f**). Dashed lines indicate the expected relationship if RDT and PCR detected equal prevalence. Horizontal and vertical lines around point estimates indicate 95% confidence intervals, whereas coloured solid lines indicate the median of the Bayesian posterior distributions from the fitted model and shaded areas indicate 95% credible intervals of these fits. Radius of point estimates indicate cluster size (from small to large: <50, 50–100 and >100).

DISCUSSION

As the burden of malaria continues to decline in many regions[1], it is crucial to understand the suitability of diagnostics for use in low-transmission and near-eliminating areas where MSAT and MDA strategies are likely to be applied. More specifically, how will diagnostic accuracy affect the ability of MSAT programmes to detect and treat asymptomatic individuals or determine local malaria prevalence thresholds for the initiation of MDA? Our study results show that the detection capability of RDTs is comparable with, and often greater than, microscopy. On average, microscopy captured 87% of RDT-positive infections, with higher test concordance in children than in adults. The extent to which this higher RDT detection reflects increased sensitivity, lack of specificity, or both, is unclear. Compared with molecular detection methods, however, RDTs still miss a substantial proportion of infections, capturing only 41% of PCR-positive individuals in low-transmission settings. Our analysis included cross-sectional data with paired prevalence measures by either RDT and microscopy or RDT and PCR from more than 180,000 individuals, spanning more than 400 geographical clusters. The detection levels observed differed depending on age and transmission intensity, reflecting complex dynamics at both the ecological and host level that may influence parasite densities and the relative performance of these diagnostics.

Factors correlated with the accuracy of RDTs are varied and likely to be driven by subtleties in the concentration and duration of HRP2 antigens in peripheral circulation. A lower specificity by RDT is expected given that, in addition to current infection, they can detect recent infection owing to residual HRP2 even after parasite clearance. Our analysis found that RDTs had a higher positivity rate than microscopy among those who were more likely to have current or recent high parasite densities — children, those with measured or reported fever and those recently treated with antimalarial drugs. This may indicate that high parasite densities and, therefore, ruptured schizonts (asexual parasites that replicate to form multiple red blood cell invading parasites), lead to increased and/or prolonged HRP2 levels. These levels are likely to vary depending on an individual's clinical status and stage of infection owing to associated fluctuations in parasite density. Because RDTs have been designed for clinical use, it is intuitive that their performance would be optimal in the detection of high-density infections associated with symptomatic disease. A previous analysis evaluating the sensitivity of RDTs and microscopy, specifically in

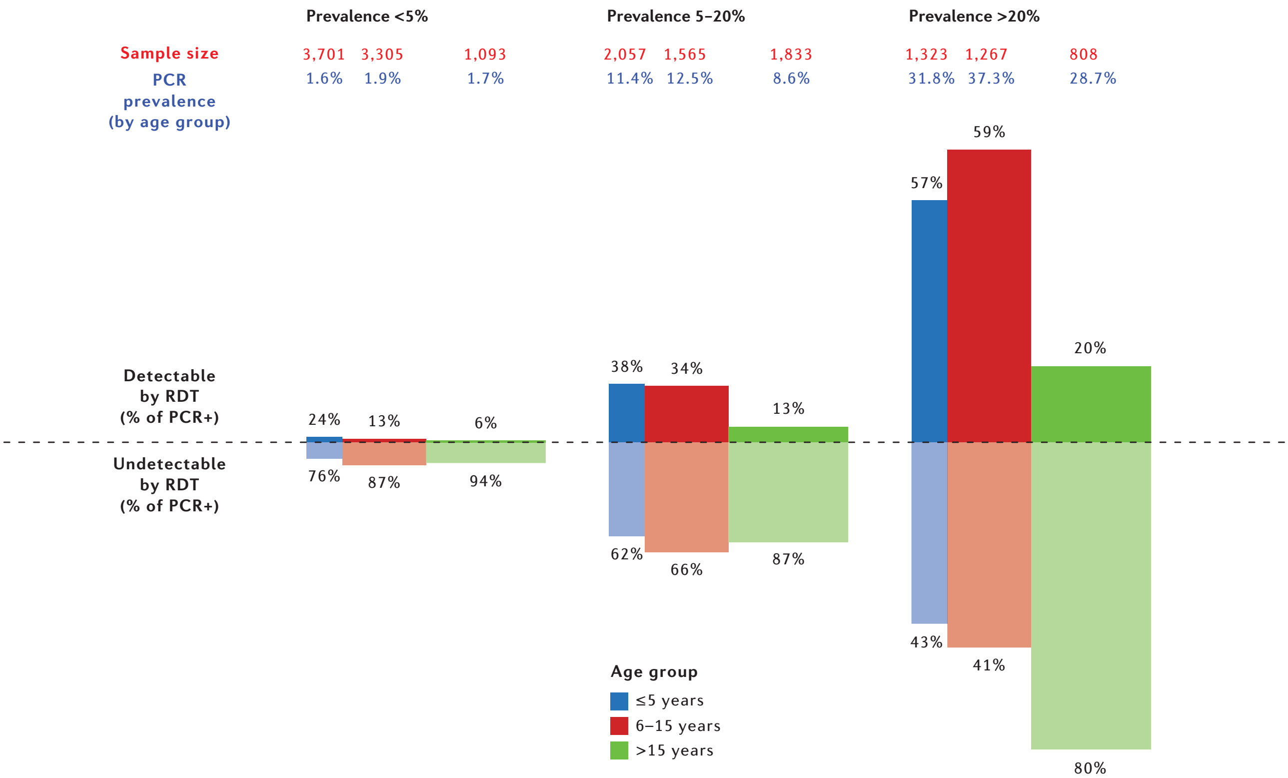


Figure 5 | Rapid diagnostic test (RDT) detectable (darker colours) and undetectable (lighter colours) infections based on polymerase chain reaction positive (PCR+) infections by age (under 5 years, 6–15 years and older than 15 years) and transmission intensity (PCR prevalence <5%, 5–20% and >20%). The height of the bars for RDT detectable and undetectable proportions reflects the total prevalence of infection in that group according to PCR, whereas the width of the bars shows the proportion of the population in each age group in most African settings (younger than 5 years (blue), 15%; 6–15 years (red), 35%; and over 15 years (green), 50% of the total population[70]).

individuals with clinical symptoms, found an association between parasite density and RDT positivity[60]. This study also stressed the issue of false positives and how RDT specificity, in addition to being influenced by parasite density, may be correlated with age and transmission intensity. Further investigation into how RDT accuracy varies between clinical and subclinical populations could help to elucidate the factors that drive these differences. Our analysis also found that using an ITN was associated with better concordance of RDT and microscopy results, most probably due to a lower risk of infection. This distinction is particularly relevant for elimination strategies, because an RDT-positive and microscopy-negative result after parasite clearance may still indicate recent transmission in a population, whereas absence of infection does not. In general, it should be noted that the quality of microscopy is likely to vary more widely than that of RDTs. Microscopy in the context of research surveys is more accurate than those typically encountered during routine surveillance[61]. Therefore, the relative sensitivity of these diagnostics may be more discordant in programmatic settings than the relationship observed in this study.

Our analysis also found a number of factors that correlated with detection by RDT and PCR. Previous studies have demonstrated that the proportion of carriers with sub-microscopic infections decreases in areas of high-transmission intensity, potentially associated with superinfection (new malaria infection in already infected individuals)[13,14]. This trend was also observed in our analysis — the proportion of PCR-measured infections that were detected by RDT increased with higher transmission intensity. Although the interaction between infection, immunity and parasite density in these settings is not fully understood, it has been suggested that only partial cross-immunity is acquired against malaria parasite clones[62]. Greater multiplicity of infection in higher transmission settings could result in higher parasite densities if host immune systems cannot respond to the diversity of parasites or if parasites increase growth rates in the presence of competing clones[14,63]. In addition to transmission intensity, we also observed age-associated variations in RDT detection. Our analysis shows that, after adjusting for transmission intensity, the odds of having an RDT-undetectable infection in adults was fivefold higher compared with under 5 year olds, potentially owing to more enhanced immune responses in adults that suppress parasite proliferation. This finding coincides well with data that show a lower sensitivity of microscopy relative to PCR among adults[13]. In addition, among PCR-positive individuals, the odds of a positive RDT result was seven times higher in patients with a fever. Overall, these results emphasize that fever, superinfections and childhood infections are commonly associated with high parasite densities, which, in turn, may lead to higher HRP2 levels that persist after parasite clearance. A number of studies have shown a relationship between parasite biomass and HRP2 clearance time[64–66]. However, these studies were predominantly in areas of high-density infections; studies in areas of lower parasite densities are less conclusive[61]. Moreover, HRP2 concentrations may be influenced by duration of infection, parasite sequestration and HRP2 antibody responses[67]. Therefore, characterizing HRP2 detection profiles at parasite densities that are more typically found in elimination settings can help to better gauge the accuracy of RDTs in these areas. Our results also showed that risk of an RDT-positive and PCR-negative test result was higher in school-aged children compared with children under 5 and adults. This may be further evidence for an association between age and recent high parasite density (approximately 2–4 weeks), but may also suggest that infections can fall below the detection limit of PCR and still be captured by RDTs. RDT results that are typically presumed to be false positives may be advantageous when the identification of a recent as well as a current infection is needed, such as in elimination settings, or if HRP2 is still measurable during periods of fluctuating parasite density that drop below the molecular detection threshold. An improved understanding of RDT performance relative to PCR methods of various sensitivities, such as qPCR and LAMP, could help to further benchmark the range at which RDTs can optimally operate. Although the impact of the PCR method on test sensitivity has been investigated in previous studies[14], more data are required to evaluate this relative to RDT sensitivity in more detail.

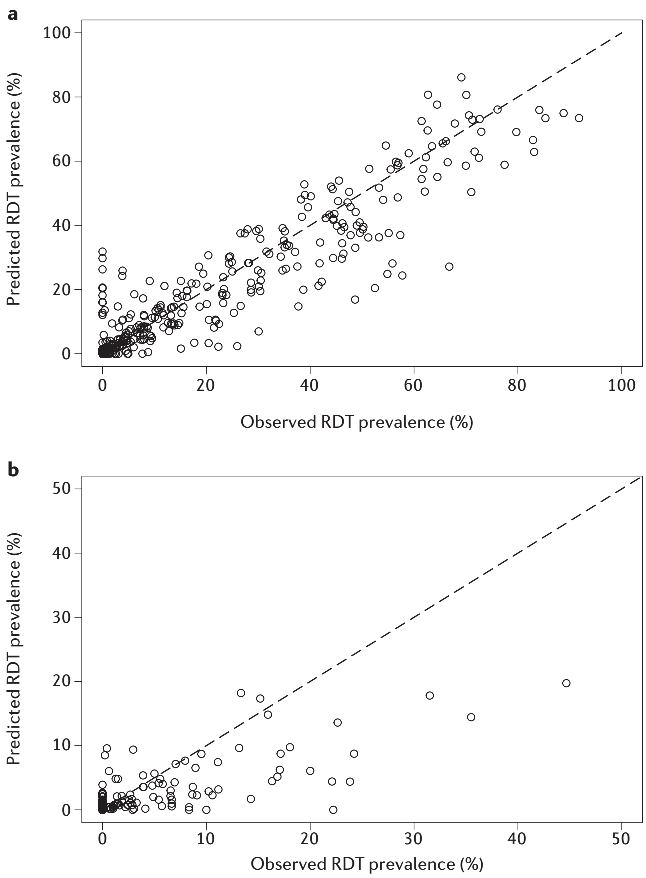


Figure 6 | The relationship between observed and predicted rapid diagnostic test (RDT) prevalence from the RDT:microscopy comparison (a), and the RDT:polymerase chain reaction comparison (b). Predictions were obtained using leave-one-out cross-validation.

We were able to define a more robust model for the relationship between prevalence measured by RDT compared with microscopy, than for the relationship between prevalence measured by RDT compared with PCR. This is because a more comprehensive data set of comparative RDT and microscopy measures was available across a wider range of transmission intensities. Medium- to high-transmission settings were particularly under-represented in the comparison of RDT and PCR measures. With more than half of our data from <5% PCR prevalence settings (57%; 93 of 162 clusters), the RDT:PCR relationship described here primarily reflects RDT performance at low-transmission intensity. However, the relationship between RDT- and PCR-measured prevalence estimated from directly observed paired data was not statistically different from the RDT:PCR relationship estimated by triangulating the RDT:microscopy and microscopy:PCR relationships based on independent data sets, improving confidence in our findings. Additional covariate information in future studies would further explain other factors that influence diagnostic sensitivity. Although we included RDT brand as a covariate in both the RDT:microscopy and RDT:PCR models, studies in this meta-analysis were not collected specifically to evaluate RDT brand so data are not sufficiently representative to draw conclusions on its impact on diagnostic sensitivity.

Overall, this study has established the relative detection capabilities of existing diagnostics for the identification of asymptomatic individuals infected with *P. falciparum*. To inform community chemotherapy programmes, however, further analysis is needed to determine to what extent these individuals contribute to onward transmission. As with detection, the potential infectiousness of asymptomatic individuals is sensitive to fluctuations in parasite density over the course of an infection and by season[15,68]. These are driven by the maturity of the host's immune response, which may vary by age and by local transmission dynamics, such as seasonality, that can influence population-level immunity or within-host parasite behaviour. Therefore, defining infectivity in relation to parasite density is especially important; this is addressed further by Slater and colleagues in a companion paper in this supplement[69]. Once the contribution of asymptomatic individuals to the infectious reservoir is better defined, future analyses should ideally establish optimal detection limits of new diagnostics for use in control and elimination strategies.

1. World Health Organization. *World Malaria Report 2014* (WHO, 2014)
2. The malERA Consultative Group on Diagnoses and Diagnostics. A research agenda for malaria eradication: diagnoses and diagnostics. *PLoS Med* **8,** e1000396 (2011).
3. Anthony, M. P., Burrows, J. N., Duparc, S., Moehrle, J. J. & Wells, T. N. The global pipeline of new medicines for the control and elimination of malaria. *Malar. J.* **11,** 316 (2012).
4. Tietje, K. *et al.* The essential role of infection-detection technologies for malaria elimination and eradication. *Trends Parasitol.* **30,** 259–266 (2014).
5. Tiono, A. B. *et al.* Lessons learned from the use of HRP-2 based rapid diagnostic test in community-wide screening and treatment of asymptomatic carriers of *Plasmodium falciparum* in Burkina Faso. *Malar. J.* **13,** 30 (2014).
6. Cook, J. *et al.* Mass screening and treatment using a falciparum-specific rapid diagnostic test did not reduce malaria incidence in Zanzibar. *J. Infect. Dis.* **211,** jiu655 (2014).
7. Guerra, C. A. *et al.* Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project. *Malar. J.* **6,** 17 (2007).
8. Ochola, L. B., Vounatsou, P., Smith, T., Mabaso, M. L. H. & Newton, C. R. J. C. The reliability of diagnostic techniques in the diagnosis and management of malaria in the absence of a gold standard. *Lancet. Infect. Dis.* **6,** 582–588 (2006).
9. World Health Organization. *Malaria Rapid Diagnostic Test Performance. Results of WHO Product Testing of Malaria RDTs: Round 4 (2012)* (WHO, 2012).
10. Cordray, M. S. & Richards-Kortum, R. R. Emerging nucleic acid–based tests for point-of-care detection of malaria. *Am. J. Trop. Med. Hyg.* **87,** 223–230 (2012).
11. Andrews, L. *et al.* Quantitative real-time polymerase chain reaction for malaria diagnosis and its use in malaria vaccine clinical trials. *Am. J. Trop. Med. Hyg.* **73,** 191–198 (2005).
12. Rockett, R. J. *et al.* A real-time, quantitative PCR method using hydrolysis probes for the monitoring of *Plasmodium falciparum* load in experimentally infected human volunteers. *Malar. J.* **10,** 48 (2011).
13. Okell, L. C., Ghani, A. C., Lyons, E. & Drakeley, C. J. Submicroscopic infection in *Plasmodium falciparum*-endemic populations: a systematic review and meta-analysis. *J. Infect. Dis.* **200,** 1509–1517 (2009).
14. Okell, L. C. *et al.* Factors determining the occurrence of submicroscopic malaria infections and their relevance for control. *Nature Commun.* **3,** 1237 (2012).
15. Bousema, T., Okell, L., Felger, I. & Drakeley, C. Asymptomatic malaria infections: detectability, transmissibility and public health relevance. *Nature Rev. Microbiol.* **12,** 833–840 (2014).
16. Stone, W., Gonçalves, B. P., Bousema, T. & Drakeley, C. Assessing the infectious reservoir of falciparum malaria: past and future. *Trends Parasitol.* **31,** 287–296 (2015)
17. Bousema, T. & Drakeley, C. Epidemiology and infectivity of *Plasmodium falciparum* and *Plasmodium vivax* gametocytes in relation to malaria control and elimination. *Clin. Microbiol. Rev.* **24,** 377–410 (2011).
18. Schneider, P. *et al.* Quantification of *Plasmodium falciparum* gametocytes in differential stages of development by quantitative nucleic acid sequence-based amplification. *Mol. Biochem. Parasitol.* **137,** 35–41 (2004).
19. Dondorp, A. M. *et al.* Estimation of the total parasite biomass in acute falciparum malaria from plasma PfHRP2. *PLoS Med* **2,** e204 (2005).
20. McMorrow, M. L., Aidoo, M. & Kachur, S. P. Malaria rapid diagnostic tests in elimination settings--can they find the last parasite? *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **17,** 1624–1631 (2011).
21. Florey, L. *Measures of Malaria Parasitemia Prevalence in National Surveys: Agreement Between Rapid Diagnostic Tests and Microscopy.* DHS Analytical Studies No. 43. (2014).
22. Mouatcho, J. C. & Goldring, J. P. D. Malaria rapid diagnostic tests: challenges and prospects. *J. Med. Microbiol.* **62,** 1491–1505 (2013).
23. Endeshaw, T. *et al.* Evaluation of light microscopy and rapid diagnostic test for the detection of malaria under operational field conditions: a household survey in Ethiopia. *Malar. J.* **7,** 118 (2008).
24. Falade, C. O. *et al.* Blood banking in a malaria-endemic area: evaluating the problem posed by malarial parasitaemias. *Ann. Trop. Med. Parasitol.* **103,** 383–392 (2009).
25. Ganguly, S. *et al.* High prevalence of asymptomatic malaria in a tribal population in Eastern India. *J. Clin. Microbiol.* **51,** 1439–1444 (2013).
26. Gitonga, C. W. *et al.* Use of rapid diagnostic tests in malaria school surveys in Kenya: does their under-performance matter for planning malaria control? *Am. J. Trop. Med. Hyg.* **87,** 1004–1011 (2012).
27. Golassa, L., Enweji, N., Erko, B., Aseffa, A. & Swedberg, G. Detection of a substantial number of sub-microscopic *Plasmodium falciparum* infections by polymerase chain reaction: a potential threat to malaria control and diagnosis in Ethiopia. *Malar. J.* **12,** 352 (2013).
28. Gonçalves, L. *et al.* Bayesian latent class models in malaria diagnosis. *PLoS ONE* **7,** e40633 (2012).
29. Ishengoma, D. S. *et al.* Accuracy of malaria rapid diagnostic tests in community studies and their impact on treatment of malaria in an area with declining malaria burden in north-eastern Tanzania. *Malar. J.* **10,** 176 (2011).
30. Keating, J., Miller, J. M., Bennett, A., Moonga, H. B. & Eisele, T. P. *Plasmodium falciparum* parasite infection prevalence from a household survey in Zambia using microscopy and a rapid diagnostic test: implications for monitoring and evaluation. *Acta Trop.* **112,** 277–282 (2009).

31. Laurent, A. *et al.* Performance of HRP-2 based rapid diagnostic test for malaria and its variation with age in an area of intense malaria transmission in southern Tanzania. *Malar. J.* **9,** 294 (2010).
32. Mboera, L. E. G. *et al.* Comparison of the Paracheck-Pf test with microscopy, for the confirmation of *Plasmodium falciparum* malaria in Tanzania. *Ann. Trop. Med. Parasitol.* **100,** 115–122 (2006).
33. Neumann, C. G. *et al.* Comparison of blood smear microscopy to a rapid diagnostic test for *in-vitro* testing for *P. falciparum* malaria in Kenyan school children. *East Afr. Med. J.* **85,** 544–549 (2008).
34. Satoguina, J. *et al.* Comparison of surveillance methods applied to a situation of low malaria prevalence at rural sites in The Gambia and Guinea Bissau. *Malar. J.* **8,** 274 (2009).
35. Shekalaghe, S. A. *et al.* Submicroscopic *Plasmodium falciparum* gametocyte carriage is common in an area of low and seasonal transmission in Tanzania. *Trop. Med. Int. Heal.* **12,** 547–553 (2007).
36. Sousa-Figueiredo, J. C. *et al.* Investigating portable fluorescent microscopy (CyScope®) as an alternative rapid diagnostic test for malaria in children and women of child-bearing age. *Malar. J.* **9,** 245 (2010).
37. Wanji, S., Kimbi, H. K., Eyong, J. E., Tendongfor, N. & Ndamukong, J. L. Performance and usefulness of the Hexagon rapid diagnostic test in children with asymptomatic malaria living in the Mount Cameroon region. *Malar. J.* **7,** 89 (2008).
38. Ye, Y., Madise, N., Ndugwa, R., Ochola, S. & Snow, R. W. Fever treatment in the absence of malaria transmission in an urban informal settlement in Nairobi, Kenya. *Malar. J.* **8,** 160 (2009).
39. Dal-Bianco, M. P. *et al.* High prevalence of asymptomatic *Plasmodium falciparum* infection in Gabonese adults. *Am. J. Trop. Med. Hyg.* **77,** 939–942 (2007).
40. Fancony, C., Sebastiao, Y., Pires, J., Gamboa, D. & Nery, S. Performance of microscopy and RDTs in the context of a malaria prevalence survey in Angola: a comparison using PCR as the gold standard. *Malar. J.* **12,** 284 (2013).
41. Ouattara, A. *et al. Plasmodium falciparum* infection and clinical indicators in relation to net coverage in central Cote d'Ivoire. *Parasites Vectors* **7,** 306 (2014).
42. Faucher, J.-F. *et al.* What would PCR assessment change in the management of fevers in a malaria endemic area? A school-based study in Benin in children with and without fever. *Malar. J.* **9,** 224 (2010).
43. ICF International, C. M. *Demographic and Health Surveys* (ICF, 2012).
44. Yeka, A. *et al.* Factors associated with malaria parasitemia, anemia and serological responses in a spectrum of epidemiological settings in Uganda. *PLoS ONE* **10,** e0118901 (2015).
45. Mwingira, F., Genton, B., Kabanywanyi, A.-N. M. & Felger, I. Comparison of detection methods to estimate asexual *Plasmodium falciparum* parasite prevalence and gametocyte carriage in a community survey in Tanzania. *Malar. J.* **13,** (2014).
46. Harris, I. *et al.* A large proportion of asymptomatic *Plasmodium* infections with low and sub-microscopic parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for malaria diagnostics in an elimination setting. *Malar. J.* **9,** 254 (2010).
47. Mharakurwa, S. *et al.* Pre-amplification methods for tracking low-grade *Plasmodium falciparum* populations during scaled-up interventions in Southern Zambia. *Malar. J.* **13,** 89 (2014).
48. Yilmaz, B. *et al.* Gut microbiota elicits a protective immune response against malaria transmission. *Cell* **159,** 1277–1289 (2014).
49. Proietti, C. *et al.* Influence of infection on malaria-specific antibody dynamics in a cohort exposed to intense malaria transmission in northern Uganda. *Parasite Immunol.* **35,** 164–173 (2013).
50. Stevenson, J. C. *et al.* Reliability of school surveys in estimating geographic variation in malaria transmission in the Western Kenyan Highlands. *PLoS ONE* **8,** e77641 (2013).
51. Aydin-Schmidt, B. *et al.* Loop mediated isothermal amplification (LAMP) accurately detects malaria DNA from filter paper blood samples of low density parasitaemias. *PLoS One* **9,** e103905 (2014).
52. Brown, T. *et al.* Molecular surveillance for drug-resistant *Plasmodium falciparum* in clinical and subclinical populations from three border regions of Burma/Myanmar: cross-sectional data and a systematic review of resistance studies. *Malar. J.* **11,** 333 (2012).
53. Cook, J. *et al.* Loop-mediated isothermal amplification (LAMP) for point-of-care detection of asymptomatic low-density malaria parasite carriers in Zanzibar. *Malar. J.* **14,** 43 (2015).
54. Stauffer, W. M. *et al.* Evaluation of malaria screening in newly arrived refugees to the United States by microscopy and rapid antigen capture enzyme assay. *Pediatr. Infect. Dis. J.* **25,** 948–950 (2006).
55. Stresman, G. H. *et al.* A method of active case detection to target reservoirs of asymptomatic malaria and gametocyte carriers in a rural area in Southern Province, Zambia. *Malar. J.* **9,** 265 (2010).
56. Sharp, S. J. & Thompson, S. G. Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. *Stat. Med.* **19,** 3251–3274 (2000).
57. Hay, S. I., Smith, D. L. & Snow, R. W. Measuring malaria endemicity from intense to interrupted transmission. *Lancet. Infect. Dis.* **8,** 369–378 (2008).
58. World Health Organization. *From Malaria Control to Malaria Elimination. A Manual for Elimination Scenario Planning* (WHO, 2014).
59. Snounou, G. *et al.* High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. *Mol. Biochem. Parasitol.* **61,** 315–320 (1993).
60. Abeku, T. A. *et al.* Determinants of the accuracy of rapid diagnostic tests in malaria case management: evidence from low and moderate transmission settings in the East African highlands. *Malar. J.* **7,** (2008).
61. World Health Organisation. *Parasitological Confirmation of Malaria Diagnosis. Report of a WHO Technical Consultation* (WHO, 2010).
62. Ofosu-Okyere, A. *et al.* Novel *Plasmodium falciparum* clones and rising clone multiplicities are associated with the increase in malaria morbidity in Ghanaian children during the transition into the high transmission season. *Parasitology* **123,** 113–23 (2001).
63. Pollitt, L. C. *et al.* Competition and the evolution of reproductive restraint in malaria parasites. *Am. Nat.* **177,** 358–67 (2011).
64. Kyabayinze, D. J., Tibenderana, J. K., Odong, G. W., Rwakimari, J. B. & Counihan, H. Operational accuracy and comparative persistent antigenicity of HRP2 rapid diagnostic tests for *Plasmodium falciparum* malaria in a hyperendemic region of Uganda. *Malar. J.* **7,** 221 (2008).
65. Swarthout, T. D., Counihan, H., Senga, R. K. & Broek, I. van den. Paracheck-Pf® accuracy and recently treated *Plasmodium falciparum* infections: is there a risk of over-diagnosis? *Malar. J.* **6,** 58 (2007).
66. Houzé, S., Boly, M. D., Bras, J. Le, Deloron, P. & Faucher, J.-F. PfHRP2 and PfLDH antigen detection for monitoring the efficacy of artemisinin-based combination therapy (ACT) in the treatment of uncomplicated falciparum malaria. *Malar. J.* **8,** 211 (2009).
67. Aydin-Schmidt, B. *et al.* Usefulness of *Plasmodium falciparum*-specific rapid diagnostic tests for assessment of parasite clearance and detection of recurrent infections after artemisinin-based combination therapy. *Malar. J.* **12,** 349 (2013).
68. Tusting, L. S., Bousema, T., Smith, D. L. & Drakeley, C. Measuring changes in *Plasmodium falciparum* transmission: precision, accuracy and costs of metrics. *Adv. Parasitol.* **84,** 151–208 (2014).
69. Slater, H. *et al.* Assessing the impact of next-generationrapid diagnostic testson *Plasmodium falciparum* malaria elimination strategies *Nature* **528,** S94–S101 (2015).
70. United Nations, Department of Economic & Social Affairs, Population Division. *World Population Prospects, the 2010 Revision* (UN, 2010).

SUPPLEMENTARY MATERIAL

Is linked to the online version of this paper at: http://dx.doi.org/10.1038/nature16039

ACKNOWLEDGEMENTS

We thank F. Mwingira and I. Felger (Tanzania), I. Harris and Q. Cheng (Solomon Islands), B. Yilmaz, S. Portugal and M. Soares (Mali), S. Mharakurwa and S. Volkman (Zambia), S. Staedke and G. Dorsey (Uganda), C. Proietti and T. Bousema (Uganda), J. Stevenson, G. Stresman and J. Cox (Kenya), H. Kafy, A. Bashir, E. Malik, A. Mnzava, K. Subramaniam, K. Elmardi, I. Kleinschmidt and M. Donnelly (Sudan), M. Al-Selwei, S. Al-Eryani, A. Mnzava, A. Al-Samei, K. Mustafa, H. Atta, H. Al-Yarie, G. Zamanai and C. Barwa (Yemen) for kindly sharing their data from prevalence surveys, and the Demographic and Health Survey Programme for providing survey data. We would also like to thank L. Grignard, N. Alexander and J. Cook for sharing their expertise on statistical and diagnostic methods. This study was funded by the Bill and Melinda Gates Foundation (BMGF) Diagnostics Modelling Consortium. P.W. and L.O. acknowledge funding from fellowships jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat Agreement. L.W. acknowledges doctoral training funding from the UK MRC. A.G. acknowledges support from the BMGF, the Medicines for Malaria Venture, the UK MRC and the UK DFID. C.D. is funded by the Wellcome Trust grant number 091924.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests. Financial support for this publication has been provided by the Bill & Melinda Gates Foundation.

ADDITIONAL INFORMATION

ARTICLE OPEN

Assessing the impact of next-generation rapid diagnostic tests on *Plasmodium falciparum* malaria elimination strategies

Hannah C. Slater*[1], Amanda Ross*[2,3], André Lin Ouédraogo[4,5], Lisa J. White[6,7], Chea Nguon[8], Patrick G.T. Walker[1], Pengby Ngor[6,8], Ricardo Aguas[6], Sheetal P. Silal[9], Arjen M. Dondorp[6,7], Paul La Barre[10], Robert Burton[10], Robert W. Sauerwein[11], Chris Drakeley[12], Thomas A. Smith[2,3], Teun Bousema[11,12] & Azra C. Ghani[1]

Mass-screen-and-treat and targeted mass-drug-administration strategies are being considered as a means to interrupt transmission of *Plasmodium falciparum* malaria. However, the effectiveness of such strategies will depend on the extent to which current and future diagnostics are able to detect those individuals who are infectious to mosquitoes. We estimate the relationship between parasite density and onward infectivity using sensitive quantitative parasite diagnostics and mosquito feeding assays from Burkina Faso. We find that a diagnostic with a lower detection limit of 200 parasites per microlitre would detect 55% of the infectious reservoir (the combined infectivity to mosquitoes of the whole population weighted by how often each individual is bitten) whereas a test with a limit of 20 parasites per microlitre would detect 83% and 2 parasites per microlitre would detect 95% of the infectious reservoir. Using mathematical models, we show that increasing the diagnostic sensitivity from 200 parasites per microlitre (equivalent to microscopy or current rapid diagnostic tests) to 2 parasites per microlitre would increase the number of regions where transmission could be interrupted with a mass-screen-and-treat programme from an entomological inoculation rate below 1 to one of up to 4. The higher sensitivity diagnostic could reduce the number of treatment rounds required to interrupt transmission in areas of lower prevalence. We predict that mass-screen-and-treat with a highly sensitive diagnostic is less effective than mass drug administration owing to the prophylactic protection provided to uninfected individuals by the latter approach. In low-transmission settings such as those in Southeast Asia, we find that a diagnostic tool with a sensitivity of 20 parasites per microlitre may be sufficient for targeted mass drug administration because this diagnostic is predicted to identify a similar village population prevalence compared with that currently detected using polymerase chain reaction if treatment levels are high and screening is conducted during the dry season. Along with other factors, such as coverage, choice of drug, timing of the intervention, importation of infections, and seasonality, the sensitivity of the diagnostic can play a part in increasing the chance of interrupting transmission.

Nature 528, S94–S101 (3 December 2015), DOI: 10.1038/nature16040

Plasmodium falciparum malaria was responsible for an estimated 584,000 (range 367,000–755,000) deaths in 2013, most of which occurred in young children in sub-Saharan Africa[1]. Although the burden has reduced in response to global efforts to increase the provision of proven malaria interventions such as insecticide-treated bed nets and access to health care and treatment[1], it remains high. One of the challenges in reducing malaria transmission is the long duration of infection in the human host, which in semi-immune individuals may persist for a year or more[2]. In particular, although infection often leads to disease in naive individuals, those with sufficient acquired immunity can harbour parasites — and hence be onwardly infectious to mosquitoes — without exhibiting symptoms[3]. One option for speeding the decline in transmission could be to target the asymptomatic reservoir of infection[4] by providing either periodic mass-screen-and-treat (MSAT) programmes, focal MSAT or a reactive strategy in which individuals living in the vicinity of an identified clinical case are screened and treated. However, the extent to which such strategies are able to reduce the infectious reservoir will depend on the extent to which the diagnostic used to identify infected individuals also detects those who are onwardly infectious. Another form of targeting could take place at the population level (for example a village) where mass interventions are deployed if the population prevalence

*These authors contributed equally. [1]MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, Norfolk Place, London W2 1PG, UK. [2]Swiss Tropical and Public Health Institute, Socinstrasse 57, 4002 Basel, Switzerland. [3]University of Basel, Petersplatz 1, 4001 Basel, Switzerland. [4]Institute for Disease Modelling, Bellevue, Washington 98005, USA. [5]Department of Biomedical Sciences, Centre National de Recherche et de Formation sur le Paludisme, 01 B.P. 2208, Ouagadougou, Burkina Faso. [6]Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok 10400, Thailand. [7]Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7LJ, UK. [8]National Malaria Center, Ministry of Health, Phnom Penh 12302, Cambodia. [9]Department of Statistical Sciences, University of Cape Town, Rondebosch 7701, Cape Town, South Africa. [10]PATH, 2201 Westlake Avenue, Seattle, Washington 98121, USA. [11]Radboud University Medical Center, 6525 HP Nijmegen, the Netherlands. [12]London School of Hygiene & Tropical Medicine, Keppel St, London WC1E 7HT, UK. Correspondence should be addressed to: H. C. S. e-mail hannah.slater@imperial.ac.uk.

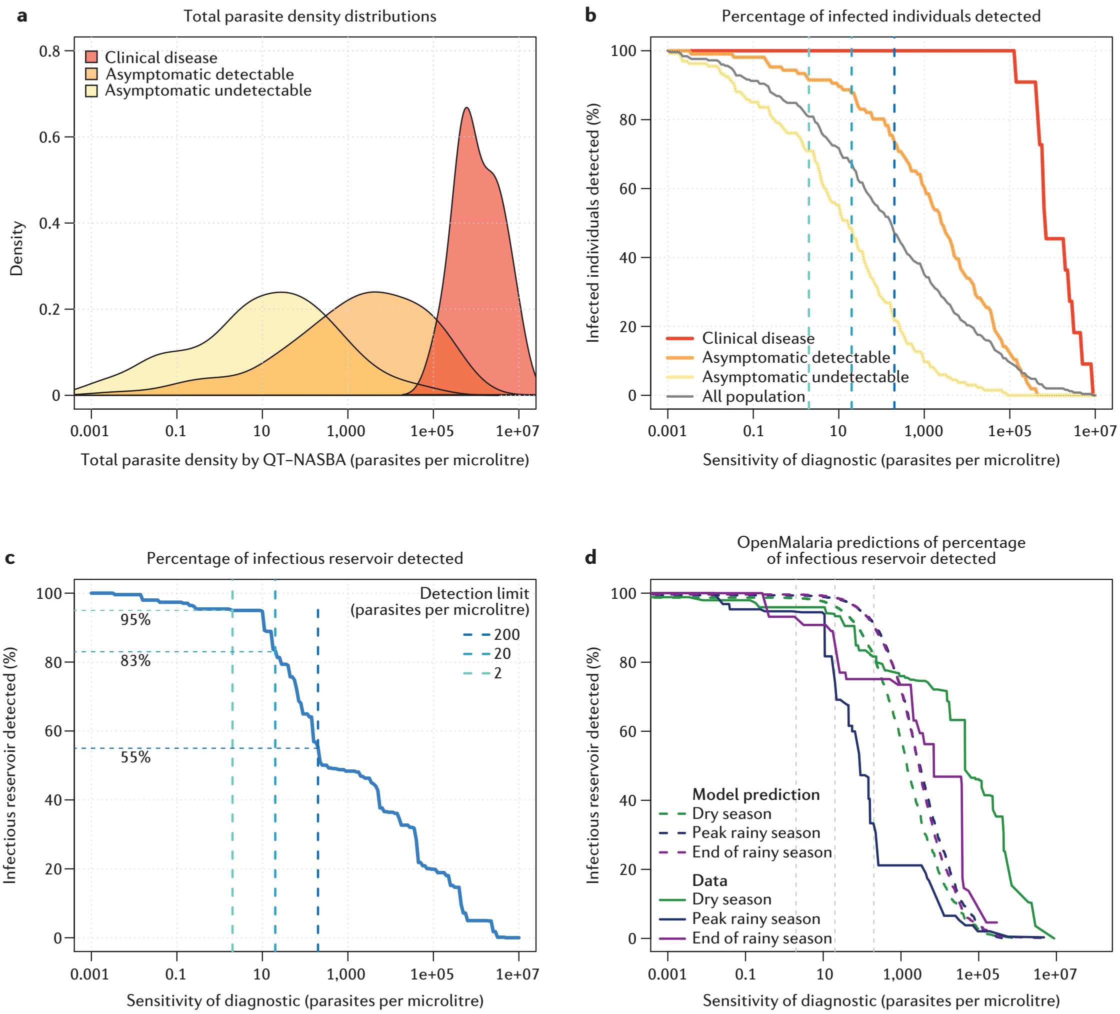


Figure 1 | Percentage of infected individuals and the infectious reservoir detected. **a**, Parasite density distributions of the three infection groups defined in the statistical analysis section. **b**, Proportion of infections detected for each infection group for a range of diagnostic thresholds between 0.001 and 10^7 parasites per microlitre. For each value of the *x*-axis we calculate the proportion of each density distribution (from **a**) that would be detected. **c**, The proportion of the infectious reservoir of the whole population that would be detected for each diagnostic threshold. This is the combined infectivity to mosquitoes of all individuals with asexual parasite densities above the diagnostic threshold weighted by body size. The dashed vertical lines show the three detection thresholds considered: 200, 20 and 2 parasites per microlitre. **d**, Proportion of the infectious reservoir detected. Comparison of the observed data as shown in **c** to a simulation-based approach using OpenMalaria. We simulated the Burkina Faso setting using OpenMalaria to see how well the distributions of parasite density and the contribution to the infectious reservoir could be captured. We assumed that the seasonality followed that reported in Burkina Faso[48] with 30 infectious bites per person per year in the village of Laye and 300 in the village of Dapélogo, and we assumed that the coverage of treatment for malaria fevers was low. The assumptions about transmission intensity and case management were not crucial to predictions in this range. We weighted each individual's contribution to the infectious reservoir by their body surface area for both the observed data and predictions to account for differential biting rates. The simulated individuals have the same age- and village-distribution as the observed data. The asexual densities were measured using quanitative nucleic acid sequence based amplification (QT-NASBA) whereas OpenMalaria output densities were calibrated to microscopy using the standard method of counting parasite against leukocytes and converting them, assuming 8,000 leukoytes per microlitre[49]. The agreement between microscopy and QT-NASBA densities is not perfect; however, we assume for the purposes of this validation that they are similar because they are two imperfect methods to measure the density of infection.

exceeds a given threshold. In this case, the choice of diagnostic will influence the accuracy of the targeting and the choice of detected prevalence threshold above which the intervention occurs.

Historically, microscopy was the most commonly used diagnostic, both in the clinic to confirm cases, and in the field to estimate the prevalence of asymptomatic infection. Although detection thresholds of 4–20 parasites per microlitre are achievable using a Giemsa-stained thick blood film in a controlled laboratory setting[5], thresholds of 100–200 parasites per microlitre are more common in field settings[6]. More recently, rapid diagnostic tests (RDTs) have been introduced in the field. These tests detect malaria antigens, most commonly histidine-rich protein 2 (HRP2) and plasmodium lactate dehydrogenase (pLDH), which are antigens produced by most, but not all, *P. falciparum* parasites[7–9]. These diagnostics are easier to use and more reliable than microscopy, but have similar detection thresholds[10,11]. The advent of highly sensitive molecular methods that detect parasite DNA or RNA such as polymerase chain reaction (PCR)[12] and quantitative nucleic acid sequence based amplification (QT-NASBA) has highlighted that many infected individuals have parasite densities that are too low to be detected either by microscopy or RDTs[13,14]. However, even these methods do not detect all the infections that are present in infected human hosts[12,15] and they are currently not available as routine field diagnostic tests because they require high-level laboratory facilities for nucleic acid extraction, amplification and detection that are unavailable in many resource-poor endemic settings. Even the most operationally attractive molecular diagnostic available — loop-mediated isothermal amplification

Table 1 | Gametocyte density, prevalence and proportion of mosquitoes infected for the three infection groups: clinical disease, asymptomatic detectable infection and asymptomatic undetectable infection, and contribution to the infectious reservoir (total infectivity weighted by body surface area).

Compartment and percentage of total infected population	Gametocyte densities (parasites per microlitre) of gametocytaemic individuals (median and interquartile range)	Proportion of individuals in compartment that are gametocytaemic (by QT-NASBA)	Percentage of mosquitoes infected by all the individuals in compartment	Contribution to infectious reservoir
Clinical disease (4.4%)	290.9 (181.7, 1,169.7)	100%	23.1%	9.0%
Asymptomatic detectable (42.2%)	89.2 (15.6, 490.1)	87%	12.7%	62.3%
Asymptomatic undetectable (53.4%)	29.8 (1.9, 128.6)	59%	3.9%	28.7%

The clinical disease group consists of the 4.4% of individuals who have the highest parasite densities. Asymptomatic detectable individuals are those that are detectable by QT-NASBA and microscopy, and the asymptomatic undetectable individuals are those detectable by QT-NASBA, but not microscopy.

(LAMP), which involves a simple DNA extraction procedure, isothermal amplification and visual examination of positivity — is currently only used in research settings[16].

Two recent MSAT trials, which used RDT screening to identify infections, failed to lead to a sustained reduction in malaria parasite prevalence or disease incidence[17,18]. It has been suggested that this could be at least partly due to ongoing transmission from low-density infections that were undetectable by the RDT or by microscopy. For example, in one MSAT study in Zanzibar, just 0.2% of people tested positive by RDT and received an antimalarial in two rounds 1 month apart, whereas 2.5% and 3.8% of people in the two rounds, respectively, were positive by PCR. RDT, therefore, detected just 8% and 5.3% of PCR-positive infections in each of the rounds, respectively[17]. A follow-up study in this setting found that LAMP was able to detect 3.4 times more infections than RDT[19]. Village-level targeted mass drug administration (MDA) is being trialled in the Greater Mekong subregion (Cambodia, Laos, Myanmar, Thailand, Vietnam and Yunnan Province in China) with a view to wider implementation[20]. In this setting a random sample of villagers is being screened using a PCR to inform decisions on MDA implementation.

To improve the impact of elimination strategies such as MSAT and targeted MDA, researchers are developing RDTs with improved limits of detection. Trying to establish the appropriate target performance specifications for these tests raises two questions. First, how infectious are individuals with low parasite densities (how much could they contribute to the infectious reservoir)? Second, could the use of a more-sensitive diagnostic increase the ability of an MSAT intervention to achieve local interruption of transmission?

We address these questions from a population perspective using a combination of data analysis and mathematical modelling. Using a data set from a study of the human infectious reservoir for malaria in a high endemic site in Burkina Faso, we estimate the contribution to the infectious reservoir of individuals with different parasite densities and hence the proportion of the infectious reservoir that could be detected across a range of diagnostic sensitivity thresholds. We then use three well-established mathematical models of malaria transmission to assess how these different diagnostic sensitivities could improve the prospect of malaria elimination in an African and an Asian context.

METHODS

Data. The data set used consists of information on 130 participants from four age groups (<5 years, 5–14 years, 15–30 years and >30 years) in the villages of Laye and Dapélogo in Burkina Faso[21]. Study participants provided 307 venous blood samples at the start of the wet season (n = 104), the peak of the wet season (n = 100) and in the subsequent dry season (n = 103). Age- and village-matched replacements were sought if individuals were lost or refused follow-up visits. Only individuals with serious acute disease, including severe clinical malaria, were excluded from participation. Finger prick blood samples were used for the preparation of microscopy slides, of which 100 microscopic fields were double read for asexual parasites and gametocytes. Nucleic acids were extracted from 100 μl venous blood samples and QT-NASBA was used to detect all parasite stages based on 18S ribosomal RNA and specifically to detect mature gametocytes based on *Pfs25* messenger RNA. The sensitivity of QT-NASBA was set to 0.01–0.02 parasites per microlitre for both 18S rRNA and *Pfs25* mRNA[22]. Parasite and gametocyte densities were estimated in relation to a standard asexual and gametocyte stage V dilution series[23]. Venous blood samples of 3 ml were fed to locally reared 4- to 5-day-old female *Anopheles gambiae sensu stricto* mosquitoes[24]; 7 days after feeding a mean of 39.3 (range 14–65) fully fed mosquitoes were dissected in 1% mercurochrome and screened for oocysts.

Statistical analysis. To estimate the infectious reservoir, we categorized the sampled individuals into four groups: clinical disease, which was defined as individuals with the highest asexual parasite densities (>140,000 parasites per microlitre); asymptomatic detectable, which was defined as individuals who were positive using QT-NASBA and microscopy; asymptomatic undetectable, which was defined as individuals who were positive by QT-NASBA and negative by microscopy; and uninfected individuals who were negative by QT-NASBA and microscopy. We define positive QT-NASBA as all individuals positive for 18S because this indicates the presence of asexual parasite stages. For the purposes of this analysis, we make the assumption that QT-NASBA is the gold standard diagnostic. Of the 1,095 mosquitoes that fed on 28 individuals in the group negative by QT-NASBA and microscopy, only 1 became infected (<0.1% of mosquitoes, 3.5% of humans), indicating that the uninfected individuals have extremely low infectiousness to mosquitoes. For each of the other groups we calculated the mean gametocyte density of gametocytaemic individuals, the proportion that were gametocytaemic using QT-NASBA, the mean gametocyte density conditional on the presence of gametocytes and the proportion of fed mosquitoes infected. The contribution of each group to the infectious reservoir was calculated as the total percentage of mosquitoes infected by each group weighted by the number of people in each group and by the relative body surface area (based on age) of those individuals. Non-parametric (Wilcoxon) tests were used to test statistical significance of differences in parasite densities between the groups. Differences in the proportions of gametocytaemic individuals and mosquitoes infected were assessed using logistic regression models using random effects to allow for repeated sampling of some of the same individuals in the different surveys.

In the models, for a given diagnostic detection limit, the proportion of the infectious reservoir that is detectable is calculated as the combined onward infectivity to mosquitoes of all individuals with parasite densities above the threshold (weighted by body surface area) divided by the infectious reservoir of the total population.

Mathematical models. Three established malaria transmission models were used to evaluate how different diagnostic thresholds would impact on the likely success of MSAT strategies. The first two models (the Imperial College model and OpenMalaria, which was developed by the Swiss Tropical and Public Health Institute and the Liverpool School of Tropical Medicine) were used to explore the impact of MSAT in an African context. The third (mathematical and economic modelling (MAEMOD) developed by the Mahidol Oxford Tropical Medicine Research Unit) was used to explore the impact of targeted MDA in an Asian context. Full mathematical details of each of the models, the process of fitting to data and their validation against further field data are given in refs 4, 25–33. We summarize how the models capture the infectious reservoir. In all three models, we define infectivity as the probability of an individual infecting a feeding mosquito, and the infectious reservoir is defined as the sum of the infectivity of the whole population multiplied by the probability of being bitten, which is based on the body surface area of the individual.

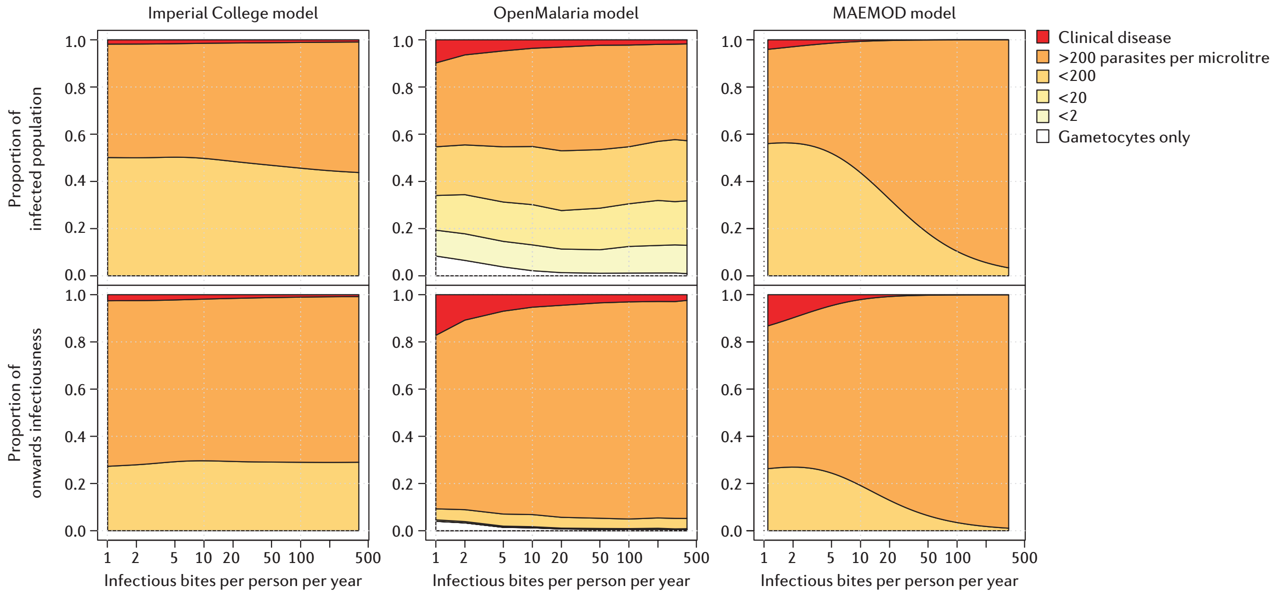


Figure 2 | Model predictions of the proportion of infected individuals in each class and their contribution to the infectious reservoir. Proportion of the infected population that have clinical disease (red), asymptomatic infection detectable by microscopy (orange) and asymptomatic infection not detectable by microscopy (yellow), and their contribution to the infectious reservoir for a range of entomological inoculation rates between 1 and 400. We predicted the proportion of infected individuals with infections detectable by microscopy for a range of transmission intensities with the three models, assuming the common inputs: stable transmission, constant seasonality, little case management and no other interventions present. The 'gametocytes only' compartment refers to individuals with recently cleared asexual parasites, but persisting gametocytes.

Imperial College

We used a deterministic age-structured population-level transmission model that tracks three infected compartments of the population — clinical disease, patent asymptomatic infection (identifiable by microscopy), and sub-patent (sub-microscopic) asymptomatic infection[29]. Onward infectivity to mosquitoes is highest in clinical disease, intermediate in patent asymptomatic infection and lowest in sub-patent infection parameterized by fitting to four earlier mosquito feeding studies[29]. As the model does not explicitly track parasite density, we adapted this structure to capture a wider range of diagnostic sensitivities by using results of the data analysis to define the proportion of infections that are detectable at a given parasite threshold within each of the infected compartments. An individual's infectivity is determined by their malaria infection status 10 days previously, to capture lags in gametocyte production.

OpenMalaria

We used an individual-based stochastic comprehensive model of malaria epidemiology[28,34]. In this model, a population of individuals is updated at 5-day time steps with model components representing new infections, parasite densities, acquired immunity, morbidity, mortality and infectivity to mosquitoes. The course of parasite densities over an infection in a non-immune individual is described using historical data from patients deliberately infected with *P. falciparum* as treatment for neurosyphilis. Immunity to asexual parasites is derived from a combination of cumulative exposure to both inoculations and parasite densities, and maternal immunity[26], and acts to reduce the densities. The model was fitted to aggregated data from several data sets but not to longitudinal patterns of densities within individuals[26]. Several different variants of this model are available. For this study we use the base model from the overall ensemble[34]. An individual's infectiousness to mosquitoes is related to a weighted sum of their recent parasite densities 10, 15 and 20 days previously, allowing time for gametocytes to develop and circulate[27]. The probability of infecting a feeding mosquito is given by the probability that at least one male and one female gametocyte would be taken up in the blood meal, and the model is fitted to the data of artificial feeding carried out on the patients with neurosyphilis treated using malaria therapy and then scaled using field data to account for the difference between infectivity in experimental and field conditions[25,27]. An individual who has recently been treated or whose infection ended naturally may have no asexual parasites, but continue to be infectious for up to 20 days. The proportion of individuals detected by a diagnostic test, and the corresponding contribution by these individuals to the infectious reservoir, is directly recorded.

MAEMOD

We used a deterministic model for the transmission of *P. falciparum* malaria similar to those previously described[32] with four infection classes: severe; clinical; asymptomatic and detectable by microscopy; and asymptomatic and undetectable by microscopy. Each infection class has a distribution of parasitaemia associated with it, which is used to estimate the sensitivity of various diagnostic tests. Each infection class also has an infectiousness associated with it based on infectivity data. We assume that treated individuals test positive for HRP2 after clearance of asexual parasitaemia for different durations, depending on the detection limit of the test.

Simulated scenarios. We considered two implementation scenarios: an MSAT programme in an African setting (Zambia), and targeted MDA in Southeast Asia (Cambodia). The former represents a setting in which malaria transmission is moderate (15% parasite prevalence measured by microscopy in children under 5 years in 2012)[35], whereas the latter is an area of low prevalence (<1% parasite prevalence across all ages measured by microscopy in 2010)[36]. Three diagnostic detection thresholds were considered: 200, 20 and 2 parasites per microlitre. The highest level was chosen to reflect the diagnostic threshold for microscopy and widely used rapid diagnostic tests, although more modern RDTs may have a sensitivity closer to 40 parasites per microlitre. The lower thresholds were chosen to reflect log orders of difference, while remaining a feasible target product profile for a new diagnostic test based on current technologies. A diagnostic threshold of 2 parasites per microlitre is currently achieved by LAMP and nested PCR[12].

MSAT in an African setting

The frequency, coverage and artemisinin-based combination therapy used were based on the current operational strategy being implemented in Zambia. A single seasonality profile was used based on average rainfall patterns in Zambia between 2002 and 2009 obtained from the US Climate Prediction Center[37]. We assumed three rounds of MSAT were conducted 1 month apart during the dry season and repeated annually for a maximum of 8 years.

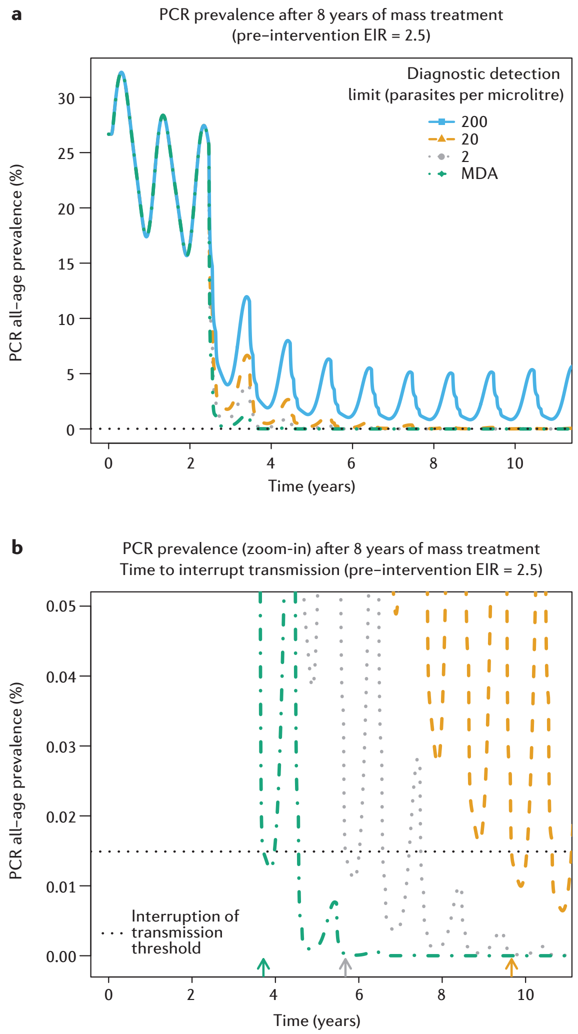


Figure 3 | Example simulation of time to interrupt transmission. a, Time series plots from the Imperial College model showing the impact of 8 years of intervention that consists of three rounds of mass screen and treat (MSAT) or mass drug administration (MDA) a month apart during the dry season with dihydroartemisinin and piperaquine at a coverage of 80%. The entomological inoculation rate is assumed to be 2.5. The blue, yellow and grey lines show the impact of MSAT using a diagnostic with a sensitivity of 200, 20 and 2 parasites per microlitre. The green line shows the impact of MDA implemented with the same timing and coverage as MSAT. The dashed line shows the prevalence threshold at which, if prevalence remains below this line for 50 consecutive days, we claim that transmission is interrupted. **b,** Enlarged version of **a** showing the interruption of transmission prevalence threshold. The arrows show the time at which interruption of transmission is achieved and the colours of the arrows correspond to the prevalence curves. EIR, entomological inoculation rates; PCR, polymerase chain reaction.

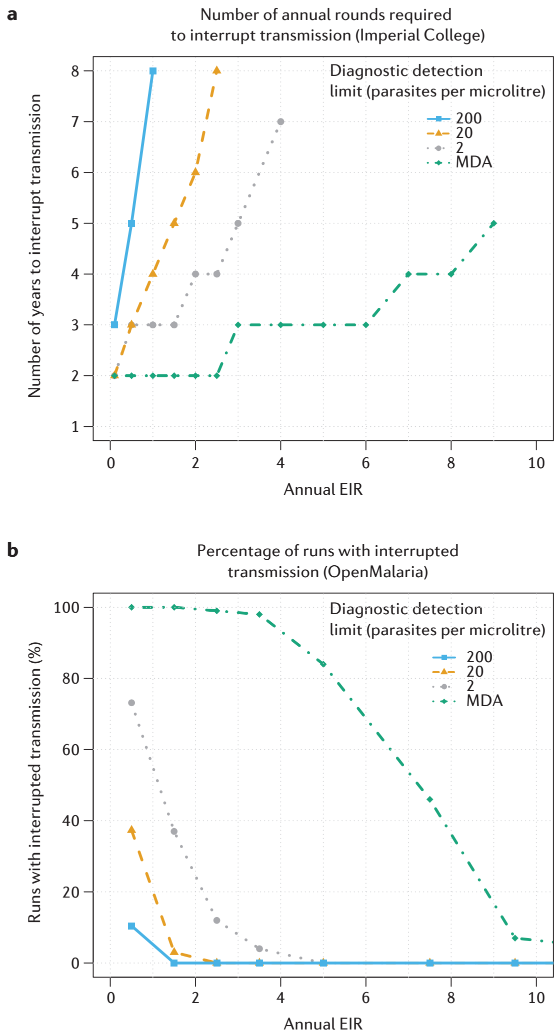


Figure 4 | Predictions of interruption of transmission for mass drug administration (MDA) and mass screen and treat (MSAT) with three diagnostic detection limits. a, Number of annual rounds required to interrupt transmission for a range of entomological inoculation rate values between 0.5 and 9.5 predicted by the Imperial College model. **b,** The probability of interrupting transmission by transmission intensity and diagnostic sensitivity predicted by OpenMalaria. We simulated scenarios for each effective transmission intensity of 10,000 individuals with no imported infections, no correlation in who receives each dose and low case-management. Because OpenMalaria is stochastic, we ran 100 simulations at each transmission intensity and present the proportion of scenarios in which transmission was interrupted. We chose one point in time to represent the results, after four years of MSAT. **a** and **b** assume three treatment rounds per year 1 month apart in the dry season.

All treatments were assumed to be with dihydroartemisinin and piperaquine, which has high efficacy in clearing parasites (95%) and a duration of protection of approximately 30 days[38]. We assumed 80% coverage of individuals at each round with no correlation between those receiving the drugs between each round, meaning that an individual has a 99% chance of being treated at least once per year. The simulations were undertaken for a range of assumed baseline transmission levels for Zambia based on entomological inoculation rates (EIRs) between 0.5 and 20 infectious bites per person per year with no imported infections from other regions .

The Imperial College model assumes that the proportion of each infected compartment that is detectable for a given diagnostic threshold is unchanged as prevalence decreases. In this model interruption of transmission is defined to have occurred once the PCR-detected prevalence for all age groups has been sustained below 0.0149% for 50 consecutive days in the absence of importation. This corresponds to a 99% probability that one or fewer people are infected in a population of 1,000 people[39]. In OpenMalaria, interruption of transmission is achieved when there are no infected individuals in a simulation. We ran 100 OpenMalaria simulations — each with a population size of 10,000 people — for each scenario to predict the probability of interrupting transmission. We did not carry out a formal model comparison.

MDA in Southeast Asia

At present, targeted MDA is considered the most promising strategy to stop the spread of artemisinin-resistant *P. falciparum* parasites in the Greater Mekong subregion[40]. In the Southeast Asian setting, targeting populations at the village level is being trialled[20]. Villages are screened to measure their

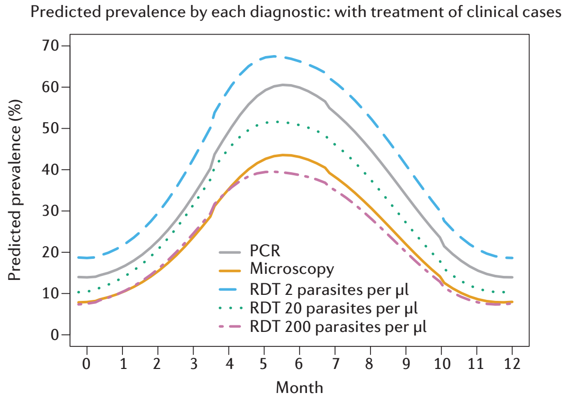


Figure 5 | The predicted seasonal pattern of prevalence by polymerase chain reaction (PCR), microscopy and rapid diagnostic test (RDT) (with detection thresholds of 200, 20 and 2 parasites per microlitre) over 1 year in a high-prevalence Cambodian setting. The results show the output from the MAEMOD model with high-coverage treatment of clinical cases.

prevalence and then if they are above a given threshold, they are selected for MDA. This is performed in the context of enhanced detection and treatment of clinical cases and high-coverage vector control. Currently, the diagnostic used is PCR. We used the MAEMOD model to investigate how using an RDT with a sensitivity of 200, 20 or 2 parasites per microlitre would affect the identification of villages. Simulations were performed for a single-peaked high-transmission Cambodian setting[41] where a village (with a population of roughly 1,000 individuals) can have a prevalence as high as 70%. This prevalence would correspond to an EIR of 3.5 per year (assuming little or no treatment of clinical cases) or 4.5 per year (assuming high coverage of treatment of clinical cases) and single-peaked seasonal transmission. The predicted prevalence in a village over time with the different diagnostics was used to assess the ability of the diagnostic to select a village for targeted MDA.

RESULTS

Table 1 summarizes the infection status and infectivity of the 128 participants who were sampled once, twice or three times over one annual transmission cycle in Burkina Faso. Of the 277 samples, 251 were found to be positive by QT-NASBA. Overall, 4% (11) of infections were considered to have clinical disease (associated with high parasite density), 42% (106) were considered to have asymptomatic detectable infections and 53% (134) were asymptomatic and negative under microscopy, but positive by QT-NASBA (referred to as asymptomatic undetectable). The remaining 26 samples were negative by both QT-NASBA and microscopy.

Those with clinical disease had the highest gametocyte and total parasite densities, whereas those with asymptomatic detectable infection had gametocyte and total parasite densities higher than those with asymptomatic undetectable infection (Table 1, Fig. 1a). The infection compartment has a significant effect on the proportion of individuals that are gametocyte positive ($P < 0.001$). This trend in parasite density and gametocytaemia is mirrored in the infectivity to mosquitoes: those with clinical disease infect a higher proportion of mosquitoes (23%) than those with microscopically asymptomatic detectable infection (13%) and those with asymptomatic undetectable infection (4%) ($P < 0.001$). In this setting, individuals with asymptomatic undetectable (sub-microscopic) infections made up 53% of the population and contributed 29% of the infectious reservoir of the population.

Under all three diagnostic thresholds considered, all individuals with clinical disease would be detected (Fig. 1b, 1c). At a diagnostic threshold of 200 parasites per microlitre, 74% of the individuals with asymptomatic detectable infection would be detected. Increasing the sensitivity of the diagnostic to 20 or 2 parasites per microlitre increases this proportion to 88% and 92%, respectively. Similarly, with a sensitivity of 200 parasites per microlitre, only 22% of individuals with asymptomatic undetectable infection would be detected. This proportion increases to 49% and 71% by increasing the sensitivity of the diagnostic to 20 or 2 parasites per microlitre, respectively. Combining these groups and weighting by relative infectivity to mosquitoes, only 55% of the infectious reservoir is detected using a threshold of 200 parasites per microlitre. We estimate that a new diagnostic with a sensitivity of 20 or 2 parasites per microlitre would be able to detect 83% and 95% of the infectious reservoir, respectively.

Figure 1d shows the proportion of the infectious reservoir detected in the Burkina Faso setting as predicted by the OpenMalaria model (the model based on parasite densities, albeit calibrated to microscopy rather than QT-NASBA densities, and therefore the model from which such predictions can be made). These generally compare well to those observed in the data (Fig. 1c,d), but with less variability with season than observed, and a slightly smaller contribution to the infectious reservoir at low parasite densities. The detectability of infections in the data was generally lowest at the peak of the transmission season, but this pattern of seasonality was not evident in OpenMalaria.

Figure 2 shows the predicted proportion of infected individuals in each infection class for a range of transmission intensities assuming no seasonality in transmission for the three models. In both the Imperial College and OpenMalaria models, these proportions are predicted to remain reasonably constant across a range of transmission intensities, although age-specific predictions show a decrease in the proportion of younger children with sub-patent infections with transmission intensity and an increase in older adults in both models (results not shown). By contrast, the MAEMOD model suggests a higher proportion of patent infection at high-transmission intensities. This difference is probably due to the data sets used for validation, with the MAEMOD model being calibrated for low-transmission settings. In both the Imperial College and MAEMOD models, at low transmission, the proportion of individuals with clinical malaria, patent infection and sub-patent infection, and the infectivity from these groups closely match those observed in the Burkina Faso data (Table 1). The OpenMalaria model suggests a similar proportion of the population in the patent and sub-patent categories (assuming a sensitivity of microscopy close to 200 parasites per microlitre), but predicts that a greater proportion of the infectious reservoir comes from those with patent infection. It also predicts a lower proportion of the infectious reservoir from sub-patent infections under the assumption of no seasonality. However, OpenMalaria fares reasonably well when the scenario inputs are tailored to the high-transmission, highly seasonal setting (Fig. 1d). Two of the models (Imperial College and OpenMalaria) were used to simulate the impact of MSAT strategies in a Zambian setting. Figure 3 shows an example of the time to interruption of transmission with a pre-intervention EIR of 2.5 using the Imperial College model. Here the model predicts MSAT using a diagnostic test with a detection limit of 200 parasites per microlitre would not achieve interruption of transmission, even after 10 years. Increasing the sensitivity of the diagnostic to 20 or 2 parasites per microlitre would achieve interruption of transmission after 10 and 6 years, respectively. The Imperial College model predicts that repeated MSAT using a diagnostic with a detection limit of 200 parasites per microlitre only interrupts transmission in areas with an EIR below 1 (Fig. 4a). Increasing the sensitivity to 20 or 2 parasites per microlitre interrupts transmission in areas with EIRs up to 2.5 and 4, respectively. Increasing the sensitivity of the diagnostic also reduces the number of rounds required to interrupt transmission. A similar pattern is predicted by the OpenMalaria model: interruption of transmission is rare using a diagnostic with a sensitivity of 200 parasites per microlitre at an EIR greater than 1 (Fig. 4b). Increasing the sensitivity to 20 or 2 parasites per microlitre increases the probability of interruption at an EIR of 1, and at a sensitivity of 2 parasites per microlitre increases the range at which interruption could occur up to an EIR of 4. Thus, in a setting with a low EIR with a seasonality pattern from Zambia, both an increase to 20 or to 2 parasites per microlitre could improve the probability of interruption and expand the range of settings in which this could be feasible. However, it should be noted that in both models MDA is predicted to be substantially more likely to interrupt transmission and potentially to increase the range of transmission intensities at which mass treatment could interrupt transmission.

Figure 5 shows the output from the MAEMOD model applied to a Cambodian setting. Owing to the persistence of HRP2 following treatment, this model reflects the seasonal variability of an RDT in predicting prevalence because when treatment levels are high, there will be a higher proportion of HRP2-positive individuals with cleared infections. This will have an impact on the ability

of the test to detect villages suitable for MDA on a pre-screening. If there are high levels of detection and treatment of clinical malaria, the model predicts that a diagnostic with a sensitivity of 200, 20 or 2 parasites per microlitre will predict a prevalence closer to the prevalence by PCR at the beginning of the season (when prevalence in increasing) than at the end of the season (when prevalence is decreasing). In this setting a threshold of 20 parasites per microlitre is predicted to be as accurate as PCR at predicting population prevalence during the early malaria season.

DISCUSSION

The success of test-and-treat strategies to clear the infectious reservoir of *P. falciparum* malaria will depend in part on the extent to which current and new diagnostic tools are able to detect individuals who are infectious to mosquitoes. Our results from a high-endemicity setting in Burkina Faso suggest that individuals with sub-microscopic infection contribute 29% to the infectious reservoir of the population. Given the close correlation between the sensitivity of RDTs and microscopy[10] it is likely that a similar proportion of the infected population would not be detected by RDTs. Thus, even if every person in the population is screened and treated at each round (which is operationally very unlikely), just under one-third of the infectious reservoir of the population would remain untreated. There is clearly a need to develop more sensitive diagnostic tools if MSAT strategies are to be successfully deployed in the field.

By stratifying the data by parasite density, we were able to relate parasite density thresholds to the proportion of the infectious reservoir captured. At a threshold of 200 parasites per microlitre, only 22% of individuals with infection that is undetectable by microscopy have parasite densities measured with QT-NASBA above this threshold. Improving the diagnostic sensitivity to 20 or 2 parasites per microlitre increases the proportion of successfully identified infected individuals to 49% and 71%, respectively. It should be noted that these results are specific to the Burkina Faso study site — an area with very high and very seasonal transmission. Further research is required to understand how the results might be modified at lower transmission, in less seasonal settings or in settings with declining transmission. Similarly, possible associations of age with the detectability and transmissibility of infections, as well as with exposure to mosquitoes, requires further study[42].

Not all infectious individuals need to be detected and treated to reduce the effective reproduction number to below 1. At low endemicity, the effective reproduction number (that takes into account the impact of all interventions) will be closer to 1 and hence the proportion of infections that need to be detected and treated will be lower than at higher endemicity. At a given frequency of MSAT rounds, there is therefore a critical threshold of infectivity that needs to be detected to reduce the reproduction number to below 1.

Using two different transmission models applied to an African setting with seasonal transmission, our results demonstrate that MSAT using a diagnostic threshold of 200 parasites per microlitre could interrupt transmission at an EIR of between 0.5 and 1 if operational coverage is high and the intervention is continued for 8 years. By increasing the sensitivity of the diagnostic tenfold this is increased to an EIR of between 0.5 and 2.5, whereas increasing the sensitivity 100-fold would expand this to an EIR of between 2 and 4. Large areas of southern and eastern Africa have EIR values below 4, indicating that repeated MSAT programmes with a high-sensitivity diagnostic test of the order of 2 parasites per microlitre could be considered as an intervention in these areas[43]. However, the EIR in Burkina Faso is much higher than 4, suggesting that MSAT would not be effective even with a highly sensitive diagnostic test (as confirmed in a recent trial[17,18]). Equally, using a higher sensitivity diagnostic could reduce the number of treatment rounds required to interrupt transmission, thus reducing the operational and financial requirements and potentially increasing the community acceptance of the intervention. These results do have a number of caveats; in particular, as shown elsewhere, there are a number of factors (including overall coverage levels, the specific drug used, the correlation of coverage between rounds[31] the degree of importation from other areas[44], the size of the programme, and the seasonality, timing and coverage of other interventions) that can have a large impact on the success of MSAT and MDA programmes. Similar impacts of diagnostic sensitivity on the probability of elimination have been found[45].

The costs and operational challenges involved in carrying out the testing may prove more important than diagnostic sensitivity in determining how many infections can be detected and treated[46]. MDA may be a better option if it is acceptable to the local population because the additional prophylactic effect that is obtained by giving the drug to the whole population means that it is always theoretically more effective than MSAT. However, conducting multiple MDA rounds may be unpopular, especially if prevalence and incidence are low. We estimate that in the Zambia-like transmission setting described, MSAT using a diagnostic with a detection limit of 2 parasites per microlitre at a coverage of 80% is equally as effective as MDA at a coverage of 65% (results not shown). In addition, the wide-scale use of MDA has been anecdotally associated with parasite resistance[47], and there are worries that it may also accelerate the spread of artemisinin- and piperaquine-resistant malaria parasites. MSAT may therefore become the preferred option if transmission cannot be interrupted with vector control alone and if diagnostics become sufficiently sensitive to detect the minimum fraction of infectious individuals to impact transmission.

Targeted MDA is being considered in the Greater Mekong subregion. The ability to accurately target villages with high numbers of infected individuals is crucial. We have used the MAEMOD model to compare the prevalence predicted by PCR, microscopy and RDT of three different detection limits. The model predicts that a 20 parasites per microlitre detection limit is as accurate as PCR at predicting population prevalence during the early malaria season and is linked to the presence of HRP2 after successful treatment (which is detectable for longer periods of time for lower detection limits). This leads to recently treated individuals testing positive by RDT after they have cleared their asexual parasites. This could be considered an advantage for measuring population prevalence for the purposes of triggering for MDA because this type of low specificity is an indication of recent infection and is therefore linked to the current true prevalence. This aspect can therefore be exploited for triggered MDA interventions, even though it would be a disadvantage for an MSAT intervention or for clinical malaria diagnosis that leads to the treatment of uninfected patients. Focal MDA could be more acceptable than MDA to local communities because pre-testing confirms that the parasite is present in the community. It could also be more attractive to policymakers, especially in Southeast Asia, as it would mean fewer courses of antimalarial drugs are given to uninfected individuals and more heavily infected communities can be identified and targeted.

QT-NASBA is the most sensitive diagnostic used in the analysis, however, it is unable to detect all infections or infectiousness, illustrated by the fact that 1 out of 1,095 mosquitoes fed on blood from 28 QT-NASBA negative individuals became infected. This may be due to degradation of nucleic acids that disproportionally affect RNA-based diagnostics[3] or to a limitation in sensitivity regardless of sample handling. Gametocytaemic individuals with sub-PCR infection were also found in a cross-sectional survey in Tanzania using a new ultra-sensitive PCR technique[12]. These findings illustrate that the available diagnostics do not detect all infections that are present in populations and that onward malaria transmission is possible (although with a low probability) from apparently parasite-negative individuals. Importantly, these sensitive molecular techniques are currently unfeasible for large-scale malaria interventions. Diagnostics in development need to strike a balance between being sensitive enough to detect a sufficient proportion of the asymptomatic reservoir, and remaining cheap, useable and having a quick turnaround of results.

1. World Health Organisation. *World Malaria Report* (WHO, 2014).
2. Bretscher, M. T. *et al.* The distribution of *Plasmodium falciparum* infection durations. *Epidemics* **3,** 109–118 (2011).
3. Bousema, T., Okell, L., Felger, I. & Drakeley, C. Asymptomatic malaria infections: detectability, transmissibility and public health relevance. *Nature Rev. Microbiol* **12,** 833–840 (2014).
4. Águas, R., White, L. J., Snow, R. W. & Gomes, M. G. M. Prospects for malaria eradication in sub-Saharan Africa. *PLoS ONE* **3,** e1767 (2008).
5. Wongsrichanalai, C., Barcus, M. J., Muth, S., Sutamihardja, A. & Wernsdorfer, W. H. A review of malaria diagnostic tools: microscopy and rapid diagnostic rest (RDT). *Am. J. Trop. Med. Hyg.* **77,** 119–127 (2007).
6. Bell, D., Wongsrichanalai, C. & Barnwell, J. W. Ensuring quality and access for malaria diagnosis: how can it be achieved? *Nature Rev. Micro.* **4,** 682–695 (2006).
7. Baker, J. *et al.* Global sequence variation in the histidine-rich proteins 2 and 3 of *Plasmodium falciparum*: implications for the performance of malaria rapid diagnostic tests. *Malar. J.* **9,** 1–12 (2010).

8. Baker, J. *et al.* Genetic diversity of *Plasmodium falciparum* histidine-rich protein 2 (PfHRP2) and its effect on the performance of PfHRP2-based rapid diagnostic tests. *J. Inf. Dis.* **192,** 870–877 (2005).
9. Cooke, A. H. *et al.* Comparison of a parasite lactate dehydrogenase-based immunochromatographic antigen detection assay (OptiMAL) with microscopy for the detection of malaria parasites in human blood samples. *Am. J. Trop. Med. Hyg.* **60,** 173–176 (1999).
10. Banoo, S. *et al.* Evaluation of diagnostic tests for infectious diseases: general principles. *Nature Rev. Micro.* S16–S28 (2008).
11. Wu, L. *et al.* Detecting *P. falciparum* infections to reduce malaria transmission – how do current diagnostics compare? *Nature* **528,** S86–S93 (2015).
12. Hofmann, N. *et al.* Ultra-sensitive detection of *Plasmodium falciparum* by amplification of multi-copy subtelomeric targets. *PLoS Med.* **12,** e1001788 (2015).
13. Okell, L. C. *et al.* Factors determining the occurrence of submicroscopic malaria infections and their relevance for control. *Nature Commun.* **3**, 1237 (2012).
14. Okell, L. C., Ghani, A. C., Lyons, E. & Drakeley, C. J. Submicroscopic infection in *Plasmodium falciparum*-endemic populations: a systematic review and meta-analysis. *J. Inf. Dis.* **200,** 1509–1517 (2009).
15. Bretscher, M. T. *et al.* Measurement of *Plasmodium falciparum* transmission intensity using serological cohort data from Indonesian schoolchildren. *Malar. J.* **12,** 21 (2013).
16. Hopkins, H. *et al.* Highly sensitive detection of malaria parasitemia in a malaria-endemic setting: performance of a new loop-mediated isothermal amplification kit in a remote clinic in Uganda. *J. Inf. Dis.* **208,** 645–652 (2013).
17. Cook, J. *et al.* Mass screening and treatment on the basis of results of a *Plasmodium falciparum*-specific rapid diagnostic test did not reduce malaria incidence in Zanzibar. *J. Inf. Dis.* **211,** 1476–1483 (2014).
18. Tiono, A. *et al.* A controlled, parallel, cluster-randomized trial of community-wide screening and treatment of asymptomatic carriers of *Plasmodium falciparum* in Burkina Faso. *Malar. J.* **12,** 79 (2013).
19. Morris, U. *et al.* Field deployment of loop-mediated isothermal amplification for centralized mass-screening of asymptomatic malaria in Zanzibar: a pre-elimination setting. *Malar. J.* **14,** 205 (2015).
20. MORU Tropical Health Network. *First Patients Enrolled in Cambodia Targeted Malaria Elimination* http://www.tropmedres.ac/first-patients-enrolled-in-cambodia-targeted-malaria-elimination (MORU, 2015).
21. Ouédraogo, A. L. *et al.* Dynamics of the human infectious reservoir for malaria determined by mosquito feeding assays and ultra-sensitive malaria diagnosis in Burkina Faso. *J. Inf. Dis.* http://dx.doi.org/ *10.1093/infdis/jiv370* (2015).
22. Schneider, P. *et al.* Quantification of *Plasmodium falciparum* gametocytes in differential stages of development by quantitative nucleic acid sequence-based amplification. *Mol. Biochem. Parasitol.* **137,** 35–41 (2004).
23. Walker, M. *et al.* Improving statistical inference on pathogen densities estimated by quantitative molecular methods: malaria gametocytaemia as a case study. *BMC Bioinformatics* **16,** 5 (2015).
24. Ouédraogo, A. L. *et al.* A protocol for membrane feeding assays to determine the infectiousness of *Plasmodium falciparum* naturally infected individuals to *Anopheles gambiae*. *Malariaworld J.* **4** (2013).
25. Killeen, G.F., Ross, A. & Smith, T. Infectiousness of malaria-endemic human populations to vectors. *Am. J. Trop. Med. Hyg.* **75,** 38–45 (2006).
26. Maire, N. *et al.* A model for natural immunity to asexual blood stages of Plasmodium falciparum malaria in endemic areas. *Am. J. Trop. Med. Hyg.* **75,** 19–31 (2006).
27. Ross, A., Killeen, G. & Smith, T. Relationships between host infectivity to mosquitoes and asexual parasite density in *Plasmodium falciparum*. *Am. J. Trop. Med. Hyg.* **75,** 32–37 (2006).
28. Smith, T. A. *et al.* Mathematical modeling of the impact of malaria vaccines on the clinical epidemiology and natural history of *Plasmodium falciparum* malaria: Overview. *Am. J. Trop. Med. Hyg.* **75,** 1–10 (2006).
29. Griffin, J. T., Ferguson, N. M. & Ghani, A. C. Estimates of the changing age-burden of *Plasmodium falciparum* malaria disease in sub-Saharan Africa. *Nature Commun.* **5,** 3136 (2014).
30. Griffin, J.T. et al. Reducing *Plasmodium falciparum* malaria transmission in Africa: a model-based evaluation of intervention strategies. *PLoS Med.* **7,** e1000324 (2010).
31. Okell, L. C. *et al.* The potential contribution of mass treatment to the control of *Plasmodium falciparum* malaria. *PLoS ONE* **6,** e20179 (2011).
32. Maude, R. *et al.* The last man standing is the most resistant: eliminating artemisinin-resistant malaria in Cambodia. *Malar. J.* **8,** 31 (2009).
33. Maude, R. J. *et al.* Optimising strategies for *Plasmodium falciparum* malaria elimination in Cambodia: primaquine, mass drug administration and artemisinin resistance. *PLoS ONE* **7**, e37166 (2012).
34. Smith, T. *et al.* Towards a comprehensive simulation model of malaria epidemiology and control. *Parasitol.* **135,** 1507–1516 (2008).
35. Central Statistical Office & Ministry of Health. *Zambia National Malaria Indicator Survey* http://www.nmcc.org.zm/files/FullReportZambiaMIS2012_July2013_withsigs2.pdf (Government of the Republic of Zambia, 2012).
36. National Centre for Parasitology, E.a.M.C. & Malaria Consortium. *Cambodia Malaria Survey, 2010* http://www.malariasurveys.org/documents/CMS%202010%20GF%20Report%20(FINAL).pdf (Phnom Penh, Cambodia, 2010).
37. National Weather Service Climate Prediction Center. *Africa Rainfall Estimates* http://www.cpc.ncep.noaa.gov/products/fews/rfe.shtml (NOAA, 2006).
38. Okell, L. C. *et al.* Contrasting benefits of different artemisinin combination therapies as first-line malaria treatments using model-based cost-effectiveness analysis. *Nature Commun.* **5,** 5606 (2014).
39. Slater, H. C., Walker, P. G. T., Bousema, T., Okell, L. C. & Ghani, A. C. The potential impact of adding ivermectin to a mass treatment intervention to reduce malaria transmission: a modelling study. *J. Inf. Dis* **210,** 1972–1980 (2014).
40. von Seidlein, L. & Dondorp, A. Fighting fire with fire: mass antimalarial drug administrations in an era of antimalarial resistance. *Expert Rev. Anti Infect. Ther.* **13,** 715–730 (2015).
41. Song, J. *et al.* Rapid and effective malaria control in Cambodia through mass administration of artemisinin-piperaquine. *Malar. J.* **9,** 57 (2010).
42. Stone, W., Gonçalves, B.P., Bousema, T. & Drakeley, C. Assessing the infectious reservoir of falciparum malaria: past and future. *Trends Parasitol.* **31,** 287–296 (2015).
43. Gething, P. *et al.* A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malar. J.* **10,** 1–16 (2011).
44. Churcher, T. S. *et al.* Measuring the path toward malaria elimination. *Science* **344,** 1230–1232 (2014).
45. Gerardin, J., Ouedraogo, A., McCarthy, K., Eckhoff, P. & Wenger, E. Characterization of the infectious reservoir of malaria with an agent-based model calibrated to age-stratified parasite densities and infectiousness. *Malar. J.* **14,** 231 (2015).
46. Silumbe, K. *et al.* Costs and cost-effectiveness of a large-scale mass testing and treatment intervention for malaria in Southern Province, Zambia. *Malar. J.* **14,** 211 (2015).
47. von Seidlein, L. & Greenwood, B. M. Mass administrations of antimalarial drugs. *Trends Parasitol.* **19,** 452–460 (2003).
48. Ouédraogo, A. L. *et al.* Seasonal patterns of *Plasmodium falciparum* gametocyte prevalence and density in a rural population of Burkina Faso. *Acta Trop.* **105,** 28–34 (2008).
49. Shute, G. in *Malaria, Principles and Practice of Malariology* (eds Wernsdorfer, W. & McGregor, I.) 781–784 (Churchill Livingstone, 1988).

ACKNOWLEDGEMENTS

The work of L.J.W. is supported by the Bill and Melinda Gates Foundation (OPP1110500) and the Wellcome-Trust Major Overseas Programme in SE Asia (106698/Z/14/Z). The work of A.R. and T.S. (Swiss TPH) is supported by Bill & Melinda Gates Foundation grant OPP1032350 and additionally the Gottfried und Julia Bangerter-Rhyner Stiftung and the Novartis Foundation for Medical Biological Research (13A13). P.B. and R.B. (PATH) gratefully acknowledges the support of the Bill & Melinda Gates Foundation (DIAMETER, OPP1053616). A.L.O. is supported by the Netherlands Organization for Higher Education in the Tropics (grant CF29132006), the Global Good Fund, Bellevue, Washington, USA and the Ministry of Health of Burkina Faso. The work of C.J.D. and T.B. is supported by the Bill & Melinda Gates Foundation (AFIRM, OPP1034789); T.B. is further supported by a fellowship of the European Research Council (ERC-2014-StG 639776). A.C.G., H.S. and P.W. acknowledge support from the Bill and Melinda Gates Foundation, the Medicines for Malaria Venture, the UK MRC and the UK Department for International Development.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests. Financial support for this publication has been provided by the Bill & Melinda Gates Foundation.

ADDITIONAL INFORMATION

ARTICLE OPEN

Health-seeking behaviour, diagnostics and transmission dynamics in the control of visceral leishmaniasis in the Indian subcontinent

Graham F. Medley[1], T. Déirdre Hollingsworth[2,3,] Piero L. Olliaro[4,5] & Emily R. Adams[2,6]

Countries in the Indian subcontinent have committed to reducing the incidence of kala-azar, a clinical manifestation of visceral leishmaniasis, to below 1 in 10,000 by 2020. We address the role of timing of use and accuracy of diagnostics in kala-azar control and elimination. We use empirical data on health-seeking behaviour and health-system performance from the Indian state of Bihar, Bangladesh and Nepal to parameterize a mathematical model. Diagnosis of cases is key to case management, control and surveillance. Treatment of cases prevents onward transmission, and we show that the differences in time to diagnosis in these three settings explain the observed differences in incidence. Shortening the time from health-care seeking to diagnosis is likely to lead to dramatic reductions in incidence in Bihar, bringing the incidence down to the levels seen in Bangladesh and Nepal. The results emphasize the importance of maintaining population and health-system awareness, particularly as transmission and disease incidence decline. We explore the possibility of diagnosing patients before the onset of clinical kala-azar (before 14 days fever), and show that this could have a marked impact on incidence, even for a moderately sensitive test. However, limited specificity (that results in false positives) is a major barrier to such a strategy. Diagnostic tests of high specificity used at an early stage of active infection, even if sensitivity is only moderate, could have a key role in the control of kala-azar, and prevent its resurgence when paired with the passive health-care system and tests of high sensitivity, such as the test for rK39 antibody response.

Nature 528, S102–S108 (3 December 2015), DOI: 10.1038/nature16042

The protozoan parasite *Leishmania* is transmitted by the bite of an infected sand fly. It disproportionately affects the poorest communities in endemic countries, and has an associated global mortality of 200,000–400,000 per year[1]. An elimination campaign has been running in the Indian subcontinent (India, Nepal, Bangladesh, Bhutan and Thailand) since 2005. Three elimination time frames currently exist; the first, ending in 2015, is to establish progress that has been made; the second is the elimination of visceral leishmaniasis as a public health problem by 2017 (committed to by Indian subcontinent governments and visceral leishmaniasis programme managers); the third, as part of the London declaration on neglected tropical diseases, is to eliminate visceral leishmaniasis as a public health problem by 2020 (defined as less than 1 case of kala-azar in 10,000 people in endemic areas, at the block (India) and upazila (sub-district; Bangladesh) level in the Indian subcontinent). In this Article we use the term kala-azar to define the clinical disease and manifestations of the infection caused by visceral leishmaniasis.

Considerable progress has been made towards the target of less than 1 case in 10,000 by implementing novel case-detection strategies, rapid diagnostic testing and vector control activities. At present, Nepal has achieved elimination for two consecutive years and Bangladesh has reached the elimination targets in all upazilas (World Health Organization (WHO), personal communication). However, India has not yet reached these low levels and the latest data are more than 1 case per 10,000 people in endemic districts. This higher rate of incidence is thought to be due to a combination of differences in underlying transmission, pre-elimination campaign endemicity, health systems, diagnosis rates and the use and success of vector control programmes[2]. However, the relative contribution of these different factors has yet to be quantified and will be a crucial determinant of the success of the expansion of control programmes.

The case-defining conditions for kala-azar in the context of the elimination programme in the Indian subcontinent are prolonged fever of more than 2 weeks, splenomegaly and a positive rK39 test. The rK39 test is an antibody-based detection, immunochromatographic test that has been shown to have high sensitivity (around 97%) when combined with clinical symptoms[3] (Table 1). There is, however, an inherent delay in receiving treatment because clinical definition of a suspect case of kala-azar requires at least 2 weeks of fever. This definition is used partly because of the low specificity of the rK39 test to identify infection rather than exposure, and to differentiate kala-azar from other fever-causing aetiologies. For the control of kala-azar, it has been proposed that a diagnostic test that detects active infection rather than the immune response to infection could be used[4]. Testing earlier could identify more patients and interrupt transmission by early treatment. However, the specificity of such a test would have to be high to avoid false positives, as many of the patients tested will present with non-specific symptoms. Consequently,

[1]Global Health and Development, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. [2]WIDER, School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK. [3]Warwick Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK. [4]UNICEF/UNDP/World Bank/WHO Special Programme on Research and Training in Tropical Diseases (TDR), World Health Organization, Geneva CH-1211, Switzerland. [5]Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, UK. [6]Research Centre for Drugs and Diagnostics, Parasitology Department, Liverpool School of Tropical Medicine, Liverpool L3 5QA, UK. Correspondence should be addressed to: E. R. A. e-mail: emily.adams@lstmed.ac.uk.

Table 1 | Current diagnostic tests with sensitivity and specificity data taken from the Indian subcontinent where possible.

Diagnostic test	Sensitivity	Specificity
rK39 RDT[3]	97.0% (95% CI = 90.0–99.5)	90.2% (95% CI = 76.1 –97.7)
DAT[21]	97.1% (95% CI = 94.9–98.4)	95.7% (95% CI = 88.1–98.5)
Parasitology spleen[22]	>95%	100%
Parasitology bone marrow[22]	60–85%	100%
Antigen (KAtex)[22]	68–100%	98%
PCR[4]	92.3% (95% CI = 88.4–94.9)	63.3 (95% CI = 53.9–71.8)

CI, confidence interval; DAT, direct agglutination test; PCR, polymerase chain reaction; RDT, rapid diagnostic test

the target product profile (TPP) of a diagnostic test that could identify less severe or even asymptomatic cases has not yet been defined. We considered the effect of such an intervention and the role of the specificity.

Within villages, cases of kala-azar tend to be clustered around index cases, suggesting that the drivers of local epidemics are individuals with kala-azar[5,6]. Most infected people are asymptomatic, but may still contribute to transmission at low levels[7]. Patients with post-kala-azar dermal leishmaniasis (PKDL) are known to harbour parasites, often in the dermal region, and are assumed to be infectious to sand flies[8]. The relative role of people with kala-azar, asymptomatic individuals and people with PKDL in sustaining transmission has not been measured directly and is unknown[6,9]. Some of this uncertainty will hopefully be addressed in the coming years, but at present more indirect methods are required to understand the dynamics of transmission and control. If kala-azar cases contribute to most of transmission, then early diagnosis and treatment is likely to be a highly effective intervention.

Very few modelling studies have been undertaken to help disentangle the interactions between individual, population and system processes for visceral leishmaniasis. This is partly because of a lack of quantitative information on the natural history of the disease[10]. The approach we present is to use a single model to compare information from three different endemic areas. From this we can infer the rates of progression through different clinical states, largely based on the fact that health seeking and health care have different outcomes in these areas. We then extend this core health-seeking model to account for transmission, using a simple, parsimonious framework to investigate the impact of changes in diagnosis on transmission dynamics.

We consider two interventions: reducing diagnostic delays in individuals who already fulfil the kala-azar definition, and introducing novel diagnostics to enable diagnosis of those who do not, or are yet to, fulfil the clinical definition of 14 days of fever and splenomegaly. We consider the dynamic consequences of these interventions, and highlight the potential for rebound epidemics as population (herd) immunity is curtailed. Finally, we consider the profile of a diagnostic required for diagnosis and treatment prior to full kala-azar, and emphasise that specificity, rather than sensitivity, is the limiting factor.

METHODS

Empirical data. The data that inform the model are from studies on self-reported time from symptoms to health seeking and eventual diagnosis of kala-azar (Fig. 1a). In Nepal and Bihar, 92 patients with kala-azar who had experienced 103 kala-azar episodes were interviewed. Patients waited for 30 days (95% confidence interval (CI) = 18–42) in Nepal before seeking health care, 3.75 times longer than in Bihar where patients waited 8 days (95% CI = 4–12). Conversely, the lag time from seeking health care to receiving a kala-azar diagnosis was 90 days (95% CI = 68–113) in Bihar compared with 25 days (95% CI = 13–38) in Nepal. The time span between diagnosis and treatment was short in both countries. In Bangladesh, a 2007 cross-sectional study in Godagari Upzila, Rajshahi, Bangladesh by the International Centre for Diarrhoeal Disease Research, of the knowledge of, attitude to, and practice surrounding kala-azar and its treatment by communities and health providers, also screened for kala-azar by rK39 dipstick test individuals who had had fever for more than 2 weeks. Around 5,000 households were surveyed, of these, 500 randomly selected household heads were interviewed, and indicated that it took 4 days

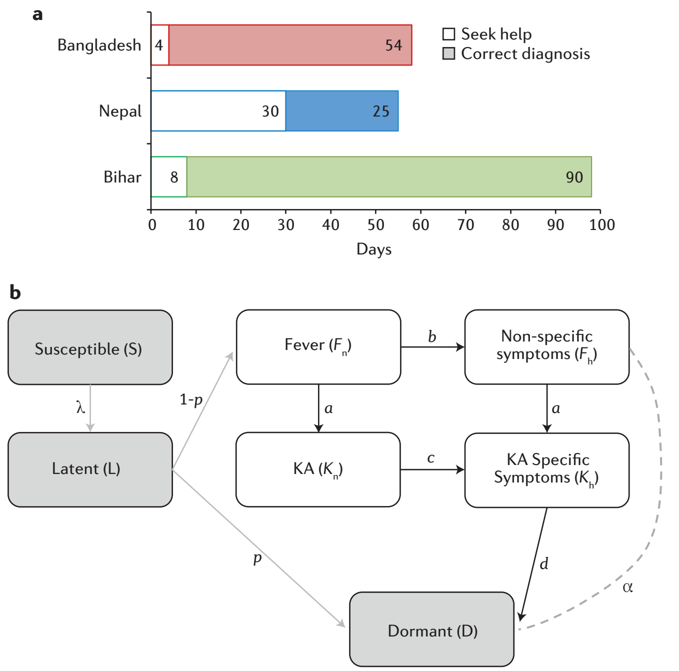


Figure 1 | **Data and model on delays in diagnosis. a,** Time from onset of symptoms (defined as fever) to health-seeking and then diagnosis, for Bihar, Nepal[14] and Bangladesh (D. Mondal, personal communication). **b,** Flow diagram of model. Open boxes show the behaviour model with progression post-infection through fever and full kala-azar (KA) (vertical flow), and from non-health-seeking to health-seeking (horizontal flow) bevaviour. The shaded boxes and grey arrows indicate the extra states required for the transmission model.

for people to seek medical help after onset of fever and 54 days until a correct diagnosis. The cumulative effect of these delays means that in Bihar patients are diagnosed 98 days after symptoms start, whereas in Bangladesh it is 58 days and in Nepal it is 55 days.

Mathematical models. We initially developed a model without transmission that mirrors the available data on the retrospective cohort of patients who have been diagnosed with kala-azar (Fig. 1b). The model includes two basic transitions: progression of disease from the point of developing fever to diagnosis, and progression to health-seeking behaviour, giving four possible states: non-health-seeking fever (F_n), health-seeking fever (F_h), non-health-seeking kala-azar (K_n) and health-seeking kala-azar (K_h). Note that we are using 'fever' to denote non-specific symptoms (the patient recognises the start of the illness that leads to a kala-azar diagnosis, but the symptoms would have been insufficiently specific for diagnosis of kala-azar at that time). In the model, given a passive health-care system and the absence of better diagnostics, only individuals who are health seeking and have clinical kala-azar can be diagnosed. We include a single parameter for disease progression (transitions from F to K, duration denoted $1/a$, where a is the rate of progression), thereby assuming that kala-azar does not have different pathology in different countries. We include three parameters that are determined by the health-seeking and diagnostic patterns and are location specific — two parameters for the onset of health seeking (F_n to F_h, and K_n to K_h rates, denoted b and c, respectively) and one parameter for diagnosis (K_h to D, denoted d). The equations are given in the Supplementary Information.

The model was parameterized to each locale. Each locale has two observations: the time from onset to health-seeking behaviour (entry into F_h or K_h), and the time from health seeking to diagnosis (K_h to D). These periods were expressed as functions of the four model parameters. Unique values of the parameters cannot be estimated, so we generated parameter sets of a, b, c and d that reproduce the observed times. We assumed that individuals with full kala-azar are more likely to seek health care than those with non-specific symptoms ($c \geq b$). A grid of all possible integer values for durations in each

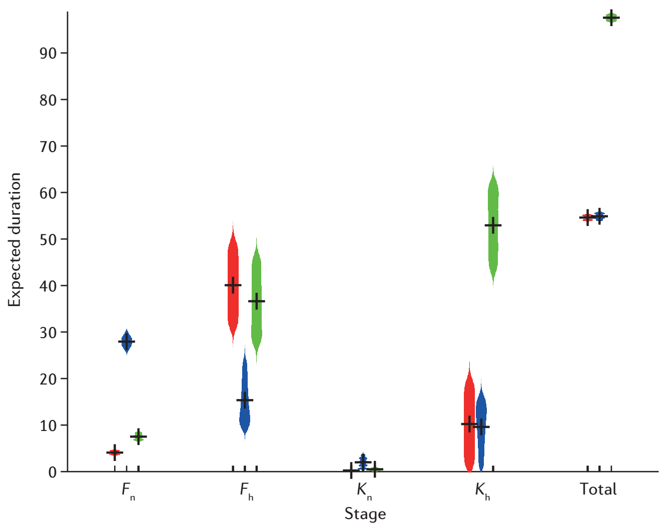


Figure 2 | Expected time in each stage of the model in each setting. The estimated duration in each stage for the combined parameter sets from each locale. From left to right, F_n, the expected time before health seeking or kala-azar (duration of time in fever, non-health-seeking state); F_h, time spent health-seeking with fever before kala-azar development; K_n, duration spent with kala-azar before health seeking; K_h, duration spent health seeking with kala-azar; total, duration spent between onset and diagnosis. Within each column, the localities are Bangladesh (red), Nepal (blue) and Bihar (green). Note that there are correlations within the parameter sets so that the total time from onset to diagnosis is constant (final column). The violin plots indicate the variability in the parameter values, and the crosses mark the mean.

state was compared with the observed outcomes. Parameter combinations that reproduce the observed times were retained, thus generating locale-specific combinations of parameters that produced results consistent with the data. The values of the parameters within the sets were correlated. For example, to reproduce the observed time to diagnosis, a shorter period spent health seeking is combined with a longer time waiting for diagnosis. Therefore, the parameters were retained as groups of appropriate parameters rather than ranges of possible values.

The model was then extended to include transmission dynamics (Fig. 1b). The extended model includes an incidence of infection and a latent class to account for delays between infection and onset of symptoms. Most latently infected individuals recover directly to the dormant stage without progressing to symptoms leading to kala-azar[11]. The instantaneous rate of infection is calculated as the weighted sum of the force of infection from the various potentially infectious states: latent, fever, kala-azar and dormant. The equations, including analytical solution for the basic reproduction number, R_0, and equilibrium state, are given in the Supplementary Information. The basic reproduction number, R_0, is defined as the average number of onward infections caused by a single infectious individual in a wholly susceptible population throughout the infectious period of that individual. R_0 is a combination of the number of infections generated in each infection stage, which in turn is determined by the duration and infectiousness of the stage. For an endemic disease such as visceral leishmaniasis, populations are not wholly susceptible, and the number of onward transmissions will be reduced by potential contacts that are already infected or immune.

The duration of the latent stage, relative infectiousness of different stages and R_0 cannot be measured with current diagnostics and available epidemiological data. Consequently, we chose values that are consistent with current understanding[10], that reproduce patterns consistent with general observations[12], and so that the equilibrium state of the model in the three settings was consistent with relative epidemiological patterns. All of the biological parameters were the same across all three settings, so differences between settings were due to the differences in health-seeking behaviour and health-care response (see Supplementary Information).

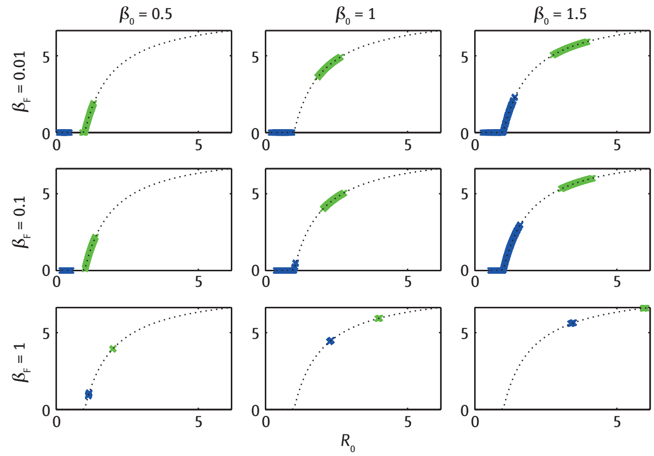


Figure 3 | Impact of diagnosis delay on transmission. Sensitivity analysis of the effect of different overall transmissibility, β_0 (columns), and the relative infectiousness of individuals with non-specific symptoms, β_F (rows). Within each panel the expected equilibrium incidence (cases per 10,000 per year) is plotted against the basic reproduction number, R_0 (dotted line). Each green point corresponds to one of the health-seeking, diagnosis and progression parameter sets for Bihar; blue are for Nepal. The points for Bangladesh are masked by those for Nepal.

We then used the model to examine two potential interventions. First, we considered the impact of reducing the time to diagnosis in Bihar to that estimated for Nepal and Bangladesh. Second, we examined the impact of a diagnostic and treatment intervention applied during the pre-kala-azar fever stage. We show results both for the average incidence at equilibrium, which may take many years to be reached, and for the average number of diagnoses over the first 5 years from introduction, to demonstrate the shorter-term effects. The impact of a diagnostic test is determined by both the sensitivity and specificity of each test, and how it is applied, for example a single test per patient or multiple testing. We include a testing rate, τ, so that the average interval between tests is $1/\tau$, and the proportion of individuals correctly diagnosed before the onset of kala-azar is $\tau S_E / (\tau S_E + a)$, where S_E is the per-test sensitivity. To account for specificity, we assume that there is a background rate of 200 additional cases of fever owing to other infections and not related to visceral leishmaniasis, and calculate the rate at which false positives arise as a function of the specificity per test, S_P, and the rate of testing. We present equilibrium results, and the numbers of true and false positives that are expected to arise over 5 years that follow the introduction of such an intervention.

RESULTS

Estimated pathways to diagnosis. The parameter sets that are consistent with the data are shown in Figure 2. There are 526 unique parameter combinations (a, b, c, d) for Bihar, 2,320 for Nepal and 312 for Bangladesh. The time between fever onset and progression to clinical kala-azar is 33 days < a < 55 days. In Nepal, individuals develop the disease faster than they seek health care, so that the typical path for an individual is fever to kala-azar to health seeking to diagnosis, with most individuals diagnosed and treated within the first few weeks of symptoms. In Bihar and Bangladesh, health seeking starts much earlier, but diagnosis is slower so that individuals enter the health system, but then remain in the symptomatic state without treatment. The major difference between Nepal and the other two regions is that kala-azar cases in Nepal most frequently first present to health-care services as clinical kala-azar, whereas in Bihar and Bangladesh people first present with non-specific symptoms. Consequently, in Bihar, patients are likely to be diagnosed and treated for more common infections (such as bacterial infection) owing to presentation of non-specific symptoms. There is a risk that treatment failure, rather than misdiagnosis, is blamed, leading to repeat treatments for the wrong diagnosis. In addition, many patients present to unqualified practitioners or the private health-care system, delaying correct diagnosis[13].

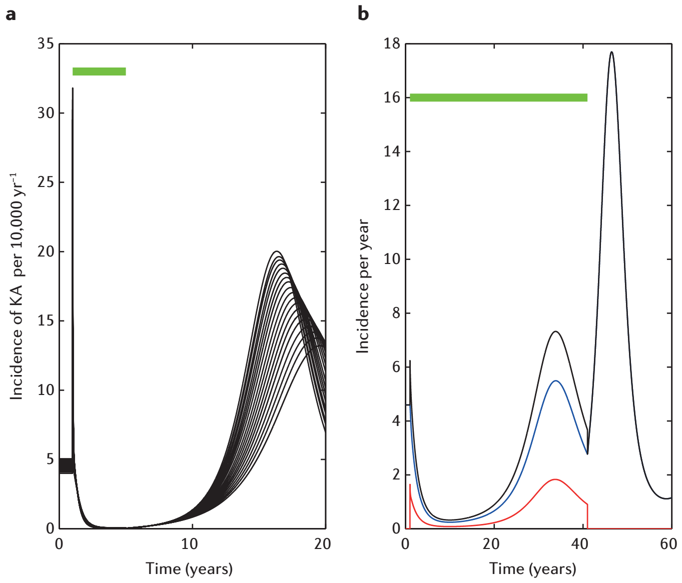


Figure 4 | The impact of interventions on transmission dynamics. a, The impact of reducing the time to diagnosis in Bihar to 10 days plus 14 days fever; the green bar indicates the duration of intervention. There is an initial increase in diagnoses of kala-azar (KA, black lines), which then leads to a reduction in underlying incidence and incidence of kala-azar. The intervention is lifted at 5 years, leading to a resurgence of cases and a large outbreak several years later. The different lines are for different parameter sets. **b,** Alternatively, the impact of introducing a diagnostic for febrile cases such that 30% of cases do not progress to kala-azar. This increases the number of diagnoses during the fever stage (red line) and decreases the number diagnosed with kala-azar (blue); the total diagnoses are shown in black. In this case, as the intervention does not lead to elimination, an epidemic occurs while the intervention is in place (green bar). When the intervention is lifted there is a further resurgence of cases.

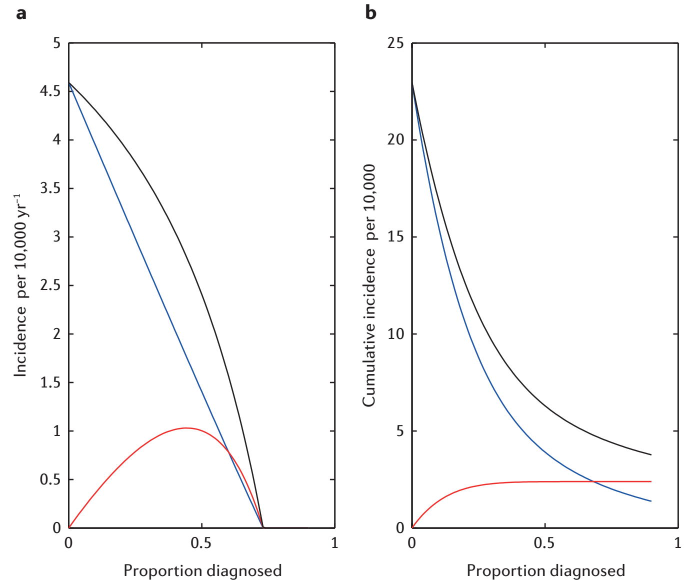


Figure 5 | The impact of a diagnostic among those with non-specific symptoms (fever). a, The equilibrium incidence of diagnosed cases (achieved after many decades) with different proportions diagnosed before full kala-azar; total incidence of diagnosed cases (black), those that are diagnosed with less specific symptoms (fever, red) and full kala-azar (blue). A 75% reduction is enough for infection elimination, 55% is enough for elimination as a public health problem (<1 case per 10,000 per year). **b,** The cumulative incidence over a shorter time scale of 5 years with different proportions diagnosed before kala-azar. Note the diminishing returns — halving the total number of cases of kala-azar can be achieved by preventing only 25% of cases from progressing.

Transmission dynamics. As with all transmission models, incidence of diagnosis increases non-linearly with the basic reproduction number, R_0, for all values of the transmission parameters (Fig. 3). At equilibrium, as long as individuals with kala-azar provide most infection to vectors, the transmission potential is higher in Bihar than Nepal (the average number of onward transmissions per infected individual, R_0, will be different). Only when transmission rates are high (Fig. 3), and therefore the number of onward infections, R_0, is also high, are the incidences of disease in all settings similar. At this level of transmission, the diagnostic differences are masked by the high infectiousness of latent, fever and dormant cases and variation between parameter sets and settings is obliterated. In reality, the transmission dynamics are not at equilibrium, and incidences are reducing in all locations due, at least in part, to reductions in R_0. Consequently, differences in incidence between settings can be explained by the differences in diagnostic delays that are consistent with the observed durations (Fig. 2). This suggests that variability in health-care seeking and diagnostic delays between different settings, and the resulting distribution of times with non-specific and specific symptoms are likely to have an impact on transmission patterns, particularly if there is differential infectivity at these different stages of infection[14].

Dynamic consequences of interventions. Owing to a lack of further information, we set the relative transmission parameters to be those in the central panel of Figure 3 (see Table 1 for parameter values). At these values, the equilibrium for Nepal and Bangladesh is low, and elimination has been achieved for some parameter sets, and the equilibrium for Bihar is between 4 and 5 cases per 10,000 people per year.

We consider two interventions that shorten the period of high infectiousness and transmission potential. The impact of these interventions on the dynamics of diagnoses is similar (Fig. 4). First, we switched the delay in diagnosis seen in Bihar (43–63 days) to the average delay in Nepal (10 days, added to the 14 days of fever to become a clinical suspect; Fig. 4a). A reduction in time to diagnosis shows an initial large peak in cases, as the cases that are 'waiting' to be diagnosed are found. This leads to a rapid reduction in transmission, which then leads to a decline in incidence. For a setting with an incidence of 5 per 10,000 people per year, elimination would be achieved with this change in the current model. Whether or not elimination would be achieved in reality depends on, among other things, the details of transmission from asymptomatic infections and spatial heterogeneities. If diagnostic delay is returned to its previous length after 4 years, then there is a rebound epidemic, but this is much slower and occurs over several years. Note that the predicted patterns are similar for all the parameter sets fitted to the Bihar situation.

Second, we introduce the diagnosis and treatment of patients while they are health seeking with non-specific febrile symptoms, before they develop kala-azar (before they have passed 14 days of fever and have splenomegaly; Fig. 4b). The dynamic response is similar to the dynamics of reducing diagnostic delays for those who have kala-azar, but without the immediate diagnostic spike. The modelled intervention (30% of kala-azar cases are diagnosed during non-specific fever) is insufficient to eliminate infection. Consequently, there is an epidemic while the intervention is in place, owing to the build-up of individuals with increased susceptibility to infection who were previously protected by the reduction in transmission. If the intervention is kept in place, then incidence eventually returns to a low level. When the intervention is removed, the second epidemic is a consequence of the increase in transmission owing to the longer infectious period when screening of fever cases is stopped.

In both interventions, the supply of full clinical kala-azar diagnoses is curtailed, transmission is reduced and there is a reduction of population (herd) immunity that leads to a bounce back of cases if the intervention is stopped. The speed at which the subsequent epidemic occurs is dependent on the success of the intervention — better curtailment of transmission results in a longer period to the next epidemic. The slow build-up means that there will be no obvious link between lengthening diagnosis delay and incidence, with clear implications for monitoring efficiency of diagnosis and treatment. In particular, if this intervention were implemented, then potentially there would be a reduction in clinical awareness of visceral leishmaniasis, resulting in a

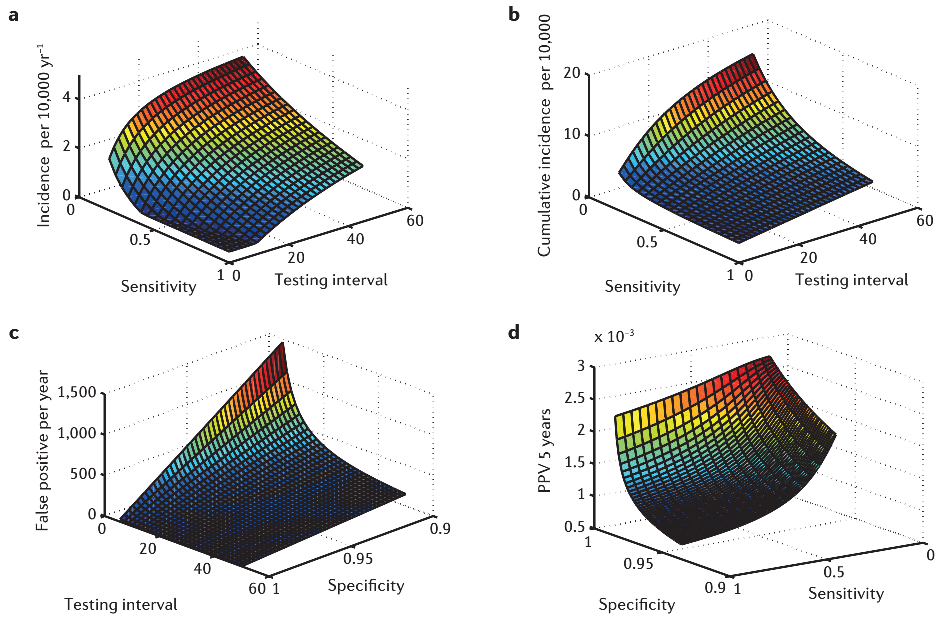


Figure 6 | The interaction between sensitivity and specificity. **a**, The equilibrium incidence of kala-azar cases as a function of the sensitivity per test and the testing interval. Elimination is only achieved for frequently applied tests (under 15 days), but can be achieved with relatively low sensitivity. **b**, As in **a** but the cumulative incidence is plotted over 5 years from the start of the intervention. This tends to reduce the importance of frequent testing for highly sensitive tests. **c**, The equilibrium incidence of false positives, assuming 200 non-visceral leishmaniasis fever cases as a function of testing interval and specificity. The relationship with specificity is linear. Note the juxtaposition of the highest numbers of false positives (short testing interval, low specificity) with the lowest numbers of true positives (short testing interval, high sensitivity in **a**). **d**, The proportion of positives that are true-positive positive predictive values, calculated as the average over 5 years after introduction of testing, as a function of sensitivity and specificity. The testing interval is fixed at 14 days (people seeking health care provide an opportunity for testing once every 2 weeks on average), and there are 200 people with non-visceral leishmaniasis fever who are health seeking. The shape is dominated by specificity owing to the numerical scales of **b** and **c**.

lengthening of diagnostic delays. We suggest that although reducing diagnosis delay through special efforts is likely to be an effective means to short-term reduction in cases, it is, depending on the epidemiological setting (baseline prevalence and biting rates), unlikely to be a sustainable route to long-term elimination. The quantitative details depend on specific parameter values and model structure, but this general pattern will be observed if those with kala-azar contribute most of the infection to vectors and onward transmission, and if there is sufficient (concomitant) immunity to kala-azar.

Impact of sensitivity of a novel diagnostic. We also consider the consequences of differing diagnostic sensitivities for diagnosis during the non-specific symptom phase in terms of its impact on equilibrium and short-term incidence of diagnoses (Fig. 5). Increased sensitivity will result in a larger proportion of cases being diagnosed earlier. At equilibrium, as the proportion of cases diagnosed increases, the incidence of diagnosis during the fever stage increases, but total diagnoses fall owing to reduced transmission (Fig. 5a). Elimination of transmission is possible even with relatively moderate sensitivity. If most transmission is from kala-azar cases, then halving the numbers progressing to this state will halve R_0. However, achieving equilibrium incidence in such models takes many decades and is mainly of theoretical interest. Consequently, we also consider the impact on the numbers of cases expected over the 5 years that follow the introduction of the intervention (Fig. 5b). Clearly, the dramatic fall in cases (Fig. 4b) occurs for low values of sensitivity, and over the short-term there is little extra to be gained from diagnosing more than 25–30% of cases before kala-azar.

Profile of the diagnostic. We profiled the diagnostic required for early case detection in terms of both sensitivity and specificity (Fig. 6). The size of the false-positive problem is determined by the frequency of fever cases (visceral leishmaniasis and non-visceral leishmaniasis) being tested for kala-azar, the specificity of the test and the proportion of fever cases that are due to visceral leishmaniasis. We, therefore, modelled a scenario in which there is a background rate of non-visceral leishmaniasis cases that present for diagnosis, and, to account for the possibility of multiple testing owing to the long duration of the fever stage for kala-azar (a known phenomenon[15]), we include a frequency of testing for health-seeking kala-azar cases with fever. If each person who seeks health care presents once, then the proportion of true cases diagnosed is the sensitivity (Fig. 5). However, if people are tested at multiple consultations, then there are multiple opportunities for a correct, true-positive result, and a lack in sensitivity can be overcome by more frequent testing (Fig. 6a,b). However, as the number of testing occasions increases, so the rate at which false positives are found increases, and increasing specificity linearly decreases the false positives (Fig. 6c). Holding the test frequency constant, the average positive predictive value (proportion of positives that are true positives) is shown in Fig. 6d as a function of sensitivity and specificity. There are likely to be many hundreds to thousands of false positives for every true positive identified unless the diagnostic, or diagnostic combination, is more than 99% specific. This problem is amplified by a decreasing prevalence of true infection, as witnessed in any elimination setting.

DISCUSSION

Our principal conclusion is that earlier diagnosis and prompt therapy have the potential to reduce ongoing transmission to elimination or near-elimination levels. This is indeed one of the pillars of the current elimination campaign, and, although this is likely to have already affected transmission in Nepal and Bangladesh, there is a large potential gain in Bihar, given that the diagnostic delays in this area are longer. We have also highlighted that curtailing transmission is likely to decrease population immunity in the long-term, so that there is a potential for large epidemics if vigilance is not maintained and diagnostic delays are allowed to increase. If diagnostic delays lengthen either the stage of health-care seeking or the ability of the system to recognise kala-azar, then subsequent epidemics are predicted to have a long lead time, and will not be immediately recognized. We have shown that the introduction of novel diagnostics on non-specific fever cases (before full kala-azar) can be effective even if sensitivity is relatively low, but that their introduction is prevented at present by less than ideal specificity, given the issues of delivery, safety and

cost of treatments. The paucity of data available to fit more complex models means that our results rely as much on understanding of infectious-disease epidemiology as on our simple model. Nonetheless, we believe that these conclusions are supported by current understanding, and, if nothing else, are valid, strong hypotheses.

We studied two mechanisms by which earlier diagnosis could occur in highly endemic settings, such as Bihar. The first is improving the health system to reduce delays in treatment. At present, differences in the time to diagnosis between countries are assumed to be because of differing time spent in the private health-care system where knowledge of kala-azar is relatively poor. Potentially, patients present with less-specific symptoms and enter a different diagnostic algorithm, and kala-azar is only suspected later. A study in Bihar using accredited social health activists (community health workers in India) to identify and refer suspected kala-azar cases showed that the time from onset of fever to seeking treatment and diagnosis at peripheral health facilities could be reduced to 32–50 days in total[16]. Our modelling reinforces the observation that reducing the time to diagnosis is an effective intervention, but as prevalence decreases it will be more difficult to maintain the necessary knowledge and infrastructure to sustain this. Intriguingly, the efficiency of the diagnostic process may have a natural equilibrium, depending on the frequency of diagnosis. We have also shown that resurgence in cases will occur after several years if transmission is reduced and not halted, and that the peak of the resurgence may well be higher than pre-intervention. Typically, local kala-azar outbreaks occur regularly — 3 peaks have occured in India 14–15 years apart over the past 40 years[17].

The second mechanism to reduce the time to diagnosis in highly endemic areas such as Bihar is to identify active infection before kala-azar onset. Specificity becomes important when designing a test that targets early case detection (before 14 days of fever). Only a test with very high specificity will allow patients to be treated, given the limited range of treatments available (toxicity, cost, administration or adherence, depending on the treatment). Nonetheless, if specificity is high enough to consider treating patients before they have 14 days of fever and become a clinical suspect, then patients could be tested and treated much earlier than is possible with the current diagnostic algorithm. This would eliminate a pool of patients that transmit visceral leishmaniasis to their local community, thereby substantially reducing future cases. Our analysis demonstrates that sensitivity of early testing for visceral leishmaniasis is not the main problem. Even a moderately sensitive test (30%) can have dramatic affect on kala-azar transmission alongside the current testing algorithm of passive surveillance with rK39. The challenge to testing earlier in the course of the disease, with the intention of treating, is avoiding a large number of false positives. However, as prevalence decreases, the positive predictive value of all tests will fall as more patients are false positives than true positives, but the resultant gain in sustainable elimination (if such a test were to be implemented) is significant.

Limitations of this model and data gaps

There are few well-conducted studies to describe the natural history of visceral leishmaniasis, which is multifactorial[18], and hence the risks for an infected individual to become diseased are not quantified. Owing to the localized epidemics that are seen surrounding index cases it is reasonable to assume that patients with kala-azar are the main reservoir of infection for sand flies. It is also thought that that those with PKDL and asymptomatic individuals or those with a dormant infection can also be infectious to sand flies, but the evidence base for this relies on anecdotal studies. Only one experimental study[8] has been published whereby sand flies were allowed to feed on four people with PKDL. Of the 400 fed flies, 104 became infected. In our analysis we have assumed that asymptomatic individuals and those with PKDL are, on average across the population, considerably less infectious than symptomatic cases[6,9]. Should asymptomatic individuals be more infectious than anticipated, the relative importance of kala-azar cases to R_0 will decrease. This means that the period of infectiousness will increase, reducing the probability of elimination and necessitating lower overall transmission rates to maintain the levels of incidence in the 1–5 cases per 10,000 people per year. If people with kala-azar are less infectious than we have assumed, the impact of reduction in diagnostic delays will be reduced, but the effect of early, novel diagnostics will be enhanced (results not shown). Of course, diagnostics for asymptomatic cases and people with PKDL may be essential for elimination if these groups are shown to be involved in transmission of disease and maintain long-term infections.

Unfortunately, without more detailed data, the models we have described are only able to highlight the qualitative patterns, although they are robust across a large range of biologically plausible parameters. The health-seeking behaviour data are informative about ranges of parameters, including the duration of non-specific symptoms, a parameter that has not been previously estimated. The data would be greatly improved by systematically sampling a population in which incidence was also estimated.

In addition to the uncertainties on the human side of transmission, we also know little of sand-fly behaviour, life expectancy and range. Clearly, these vector dynamics will have an important role in the transmission cycle and in the design of effective interventions against transmission.

Diagnostics of the future

Diagnostics play a crucial part in the control and elimination of kala-azar and are a research priority[19]. We argue that one avenue to revolutionize the current control algorithm is to develop a highly specific test of active infection (whether symptomatic or asymptomatic), even if limited in sensitivity. There is currently no suitable antigen test capable of detecting active kala-azar. The only commercially available product, KAtex, has challenges in utility and suboptimal sensitivity, although we would argue that low sensitivity should not necessarily be considered a barrier to implementation. Simplified molecular diagnostic tools that can be adapted for field situations are under development, including loop-mediated isothermal amplification[20]. These tests, along with standard polymerase chain reaction, are able to detect circulating DNA in the blood of individuals who are actively infected. Studies show that molecular tests are very sensitive, and although they have low specificity, it may be that these are infections that are below the limit of detection of the reference standard, and therefore in fact true positives[4]. Diagnostics that are able to detect asymptomatic infection and PKDL may also be important, depending on the relative role of transmission in these groups.

Given the uncertainty in clinical progression of kala-azar and diagnostic performance, there is clearly a chance that some, possibly much, of the morbidity and mortality caused by *Leishmania* infection is being misclassified. The process of diagnosis of kala-azar provides both the clinical information for treatment, as well as the data on which surveillance is built. Consequently, it is inevitable that 100% of cases of kala-azar are diagnosed and treated, regardless of the performance of the health-care system or the health-seeking behaviour of the population. Ideally, surveillance would include information that is independent of clinical diagnosis. This is particularly needed as kala-azar becomes rarer, and it is likely that both clinical and patient awareness wanes. A diagnostic tool that enables population surveillance of infection and disease, independent of clinical diagnosis, is a crucial step in achieving, enforcing and demonstrating elimination.

1. Alvar, J. *et al*. Leishmaniasis worldwide and global estimates of its incidence. *PloS ONE* **7,** e35671 (2012).
2. Mondal, D. *et al*. Visceral leishmaniasis elimination programme in India, Bangladesh, and Nepal: reshaping the case finding/case management strategy. *PLoS Neg. Trop. Dis*. **3,** e355 (2009).
3. Boelaert, M. *et al*. Rapid tests for the diagnosis of visceral leishmaniasis in patients with suspected disease. *Cochrane Database System. Rev*. **6,** CD009135 (2014).
4. de Ruiter, C. M. *et al*. Molecular tools for diagnosis of visceral leishmaniasis: systematic review and meta-analysis of diagnostic test accuracy. *J. Clin. Microbiol*. **52,** 3147–3155 (2014).
5. Bern, C., Courtenay, O. & Alvar, J. Of cattle, sand flies and men: a systematic review of risk factor analyses for South Asian visceral leishmaniasis and implications for elimination. *PLoS Negl. Trop. Dis*. **4,** e599 (2010).
6. Stauch, A. *et al*. Model-based investigations of different vector-related intervention strategies to eliminate visceral leishmaniasis on the Indian subcontinent. *PLoS Negl. Trop. Dis*. **8,** e2810 (2014).
7. Das, S., Matlashewski, G., Bhunia, G. S., Kesari, S. & Das, P. Asymptomatic *Leishmania* infections in northern India: a threat for the elimination programme? *Trans. Royal Soc. Trop. Med. Hyg*. **108,** 679–684 (2014).
8. Addy, M. & Nandy, A. Ten years of kala-azar in west Bengal, Part I. Did post-kala-azar dermal leishmaniasis initiate the outbreak in 24-Parganas? *Bull. World Health Organ*. **70,** 341–346 (1992).
9. Stauch, A. *et al*. Visceral leishmaniasis in the Indian subcontinent: modelling epidemiology and control. *PLoS Negl. Trop. Dis*. **5,** e1405 (2011).
10. Rock, K. S. *et al*. Uniting mathematics and biology for control of visceral leishmaniasis. *Trends Parasitol*. **31,** 251–259 (2015).

11. Hasker, E. *et al.* Strong association between serological status and probability of progression to clinical visceral leishmaniasis in prospective cohort studies in India and Nepal. *PLoS Negl. Trop. Dis.* **8,** e2657 (2014).
12. Dye, C. & Wolpert, D. M. Earthquakes, influenza and cycles of Indian kala-azar. *Trans. R. Soc. Trop. Med. Hyg.* **82,** 843–850 (1988).
13. Hasker, E. *et al.* Management of visceral leishmaniasis in rural primary health care services in Bihar, India. *Trop. Med. Int. Health* **15** (Suppl 2)**,** 55–62 (2010).
14. Fraser, C., Riley, S., Anderson, R. M. & Ferguson, N. M. Factors that make an infectious disease outbreak controllable. *Proc. Natl Acad. Sci. USA* **101,** 6146–6151 (2004).
15. Boettcher, J. P. *et al.* Visceral leishmaniasis diagnosis and reporting delays as an obstacle to timely response actions in Nepal and India. *BMC Infect. Dis.* **15,** 43 (2015).
16. Das, V. N. *et al.* Impact of ASHA training on active case detection of visceral leishmaniasis in Bihar, India. *PLoS Negl. Trop. Dis.* **8,** e2774 (2014).
17. Muniaraj, M. The lost hope of elimination of kala-azar (visceral leishmaniasis) by 2010 and cyclic occurrence of its outbreak in India, blame falls on vector control practices or co-infection with human immunodeficiency virus or therapeutic modalities? *Trop. Parasitol.* **4,** 10–19 (2014).
18. McCall, L. I., Zhang, W. W. & Matlashewski, G. Determinants for the development of visceral leishmaniasis disease. *PLoS Pathogens* **9,** e1003053 (2013).
19. Matlashewski, G. *et al.* Research priorities for elimination of visceral leishmaniasis. *Lancet Global Health* **2,** e683–e684 (2014).
20. Adams, E. R., Schoone, G. J., Ageed, A. F., Safi, S. E. & Schallig, H. D. Development of a reverse transcriptase loop-mediated isothermal amplification (LAMP) assay for the sensitive detection of *Leishmania* parasites in clinical samples. *Am. J. Trop. Med. Hyg.* **82,** 591–596 (2010).
21. Chappuis, F., Rijal, S., Soto, A., Menten, J. & Boelaert, M. A meta-analysis of the diagnostic performance of the direct agglutination test and rK39 dipstick for visceral leishmaniasis. *Br. Med. J.* **333,** 723 (2006).
22. Sundar, S. & Rai, M. Laboratory diagnosis of visceral leishmaniasis. *Clin. Diagnostic Lab. Immunol.* **9,** 951–958 (2002).

SUPPLEMENTARY MATERIAL

Is linked to the online version of this paper at: http://dx.doi.org/10.1038/nature16042

ACKNOWLEDGEMENTS

The authors thank D. Mondal, O. Courtenay, C. Bern, L. Chapman, A. Ghani and S. Laney for informative discussions. G.F.M. and E.R.A. gratefully acknowledge funding from the Bill and Melinda Gates Foundation through the Diagnostics Modelling Consortium based at Imperial College. G.F.M., E.R.A. and T.D.H. gratefully acknowledge funding of the NTD Modelling Consortium by the Bill and Melinda Gates Foundation in partnership with the Task Force for Global Health. The views, opinions, assumptions or any other information set out in this article are solely those of the authors. P.L.O. is a staff member of the WHO; the authors alone are responsible for the views expressed in this publication, which do not necessarily represent the decisions, policy, or views of the WHO.

COMPETING FINANCIAL INTERESTS

E.R.A. is a consultant for FIND in the evaluation of the diagnostic tool LAMP for leishmaniasis. Financial support for this publication has been provided by the Bill & Melinda Gates Foundation.

ADDITIONAL INFORMATION

ARTICLE OPEN

The role of rapid diagnostics in managing Ebola epidemics

Pierre Nouvellet*[1], Tini Garske*[1], Harriet L. Mills†[1], Gemma Nedjati-Gilani†[1], Wes Hinsley[1], Isobel M. Blake[1], Maria D. Van Kerkhove[1,2], Anne Cori[1], Ilaria Dorigatti[1], Thibaut Jombart[1], Steven Riley[1], Christophe Fraser[1], Christl A. Donnelly[1] & Neil M. Ferguson[1]

Ebola emerged in West Africa around December 2013 and swept through Guinea, Sierra Leone and Liberia, giving rise to 27,748 confirmed, probable and suspected cases reported by 29 July 2015. Case diagnoses during the epidemic have relied on polymerase chain reaction-based tests. Owing to limited laboratory capacity and local transport infrastructure, the delays from sample collection to test results being available have often been 2 days or more. Point-of-care rapid diagnostic tests offer the potential to substantially reduce these delays. We review Ebola rapid diagnostic tests approved by the World Health Organization and those currently in development. Such rapid diagnostic tests could allow early triaging of patients, thereby reducing the potential for nosocomial transmission. In addition, despite the lower test accuracy, rapid diagnostic test-based diagnosis may be beneficial in some contexts because of the reduced time spent by uninfected individuals in health-care settings where they may be at increased risk of infection; this also frees up hospital beds. We use mathematical modelling to explore the potential benefits of diagnostic testing strategies involving rapid diagnostic tests alone and in combination with polymerase chain reaction testing. Our analysis indicates that the use of rapid diagnostic tests with sensitivity and specificity comparable with those currently under development always enhances control, whether evaluated at a health-care-unit or population level. If such tests had been available throughout the recent epidemic, we estimate, for Sierra Leone, that their use in combination with confirmatory polymerase chain-reaction testing might have reduced the scale of the epidemic by over a third.

Nature 528, S109–S116 (3 December 2015), DOI: 10.1038/nature16041

The unprecedented scale of the 2014–15 West African Ebola epidemic has posed major challenges for delivering rapid diagnosis — an essential component of controlling Ebola epidemics, given the non-specific nature of early clinical symptoms. Following World Health Organization (WHO) guidelines, testing has relied on reverse transcription polymerase chain reaction (RT–PCR) based methods, which detect viral RNA in serum or plasma[1]. However, these tests are slow and costly, taking 2–6 hours to process[2] at around US$100 per test. Testing requires high levels of biosafety, and both samples and reagents must be kept cold or frozen[1,3–6]. Sustaining a cold-chain has been particularly challenging in the affected countries because of the limited infrastructure, frequent power cuts and hot climate. In addition, collecting venous blood for PCR testing requires specific medical training and poses significant risks for health-care workers[7].

Furthermore, although the laboratory processing time for PCR testing can be under 6 hours, the time between sample collection and receiving the result has often been much longer during the epidemic, owing to limited laboratory capacity and logistical infrastructure[8]. This has been a crucial issue, because delays in testing lead to longer hospital stays for patients, increasing both bed demand and the likelihood of nosocomial transmission (a major issue in Ebola outbreaks[9]). At the peak of the epidemic in Sierra Leone (between October and November 2014) there were reports of delays in test results of up to 1 week[10]. Only in January 2015, following an effort to reduce delays in laboratory access[11], did the WHO report substantially reduced average waiting times for test results of between 1.3 and 2.3 days, depending on the country[12].

Recognizing the extraordinary circumstances of the epidemic, in November 2014 WHO issued a call for "rapid, sensitive, safe and simple Ebola diagnostic tests"[2]. An ideal rapid diagnostic test (RDT) would require minimal laboratory facilities and staff training, no cold chain and could be performed with a capillary blood sample collected through a finger prick. Such an RDT could be used at the point of care to allow faster triaging of people suspected of having Ebola, returning a test result in minutes rather than days.

As a result of the WHO initiative, four tests have already been approved, including the benchmark PCR test, Realstar[13], and an RDT, ReEBOV[14–16] (Table 1). In addition, the US Food and Drug Administration has authorized ten tests for emergency use[17] and multiple alternative RDTs are in development (Supplementary Table 1).

Although recognizing that rapid and accurate diagnostics are crucial to successful containment of an Ebola outbreak, determining the best strategy and the impact of RDTs is not straightforward (Fig. 1). Poor diagnostic specificity risks patients without Ebola infection being admitted to Ebola treatment units (ETUs) — where they potentially acquire infection, and use a bed that could otherwise be used by a patient who does have Ebola. Conversely, poor diagnostic sensitivity can lead to infected individuals being discharged back to the community or sent to non-Ebola-specific wards, with the consequent risk

*These authors contributed equally. †These authors contributed equally. [1]MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, Norfolk Place, London W2 1PG, UK. [2]Center for Global Health, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France. Correspondence should be addressed to N. M. F. e-mail: neil.ferguson@imperial.ac.uk.

Table 1 | Characteristics of diagnostic tests that have been approved by the World Health Organization.

Test name	Detected	Sensitivity	Specificity	95% limit of detection	Time to result	Principal logistic challenge	Cost
RealStar Filovirus[3,13]*	Ebola-specific RNA	NR	NR	1,390 RNA cps ml^{-1} (95% CI = 690–5,320)	Hours	Kit is shipped on dry ice and should arrive frozen and be kept at −20 °C; needs equipment, including an appropriate PCR machine; needs special training; and needs high safety level in the laboratory	NR
ReEBOV Antigen Rapid Test[14-16]*	Ebola virus (EBOV) VP40 antigen	91.8% (95% CI = 84.5–96.8) compared with RealStar[15]	84.6% (95% CI = 78.8–89.4) compared to RealStar[15]	6.25E+02 ng ml^{-1} Ebola rVP40 antigen	15 minutes	Requires refrigeration at 2–8 °C, requires visual interpretation, should be used in biosafety level 4 facility or with full PPE and can use whole blood from finger prick or venipuncture	NR
Xpert[26,27]*	Ebola-specific RNA	For Ebola Mayinga RNA: 100%, (95% CI = 92.9–100.0) (PPA)	For Ebola Mayinga RNA: 100.0%, (95% CI = 92.9–100.0) (NPA)	232.4 RNA cps ml^{-1} (95% CI = 163.1–301.6)	90 minutes	Requires refrigeration at 2–8 °C, optimally should be used in a class II safety cabinet or similar, needs special training, automated process and requires a minimum of 100 μl whole blood by venipuncture	US$19.80 per cartridge
LifeRiver[4,28]	Ebola-specific RNA	1 log10 lower limit of detection than RealStar	NR	23.9 RNA cps per reaction (95% CI = 13.4–405.9)	Results in 2 hours, total processing 4–6 hours	Reagents must be kept at −20 °C; needs equipment, including an appropriate PCR machine; needs special training; needs high safety level in the laboratory	NR

*Test is also US Food and Drug Administration (FDA) authorized. All sensitivity, specificity and limits of detection are reported for Zaire EBOV unless otherwise stated. Cps, copies; NPA, negative per cent agreement; NR, not reported; PCR, polymerase chain reaction; PPA, positive per cent agreement; PPE, personal protective equipment.

of onward transmission. In this Review, we use mathematical models of Ebola transmission to explore how RDTs might best be used in future outbreaks, balancing the rapid availability of RDT results against the lower sensitivity and specificity of such tests. This trade-off makes optimization of testing strategies complex and context-dependent, from both a technical and ethical perspective.

METHODS

We examine RDTs using three different metrics, characterizing their impact on individual patient outcomes (represented by the expected case fatality ratio (CFR) for a person suspected of having Ebola who is seeking care), the effectiveness with which health-care units reduce transmission (represented by the reproduction number for a person with a true Ebola infection seeking care), and the overall scale of an epidemic (represented by the total number of cases). We evaluate the first two metrics using a model that focuses on the impact of different testing strategies implemented in the context of a single health-care unit, and the third using a model of the transmission dynamics of Ebola in Sierra Leone as a whole.

Potential impact of RDTs in a health-care unit

Figure 1 (and Supplementary Information 2) illustrates the dynamics of patients within a health-care unit represented in the health-care-unit-level model that we developed to examine three diagnostic testing strategies: PCR-only, dual testing and RDT-only. Patients are tested within the holding area or areas to determine who should be sent to a confirmed ward and who should be discharged back to the community (with v the daily number of patients seeking care and p the proportion who are infected with Ebola). During the time spent in holding, patients without Ebola have a daily risk of Ebola infection of βy, where β is the transmission rate and y is the prevalence of Ebola infection among other patients in that holding area. A similar risk affects patients misdiagnosed with Ebola in the confirmed ward. At baseline we assume $\beta = 0.15$ per day.

For analytical tractability, we model a single generation of infection: people with suspected Ebola who enter the health-care unit uninfected, but then become infected and are assumed to be discharged back to the community or sent to the confirmed ward before they become infectious. When bed capacity (the total number of beds in the health-care unit) has been exceeded by demand, we assume that patients seeking care are turned away and return to the community.

For each testing strategy we use this health-care-unit-level model to evaluate how many infected patients seeking care are sent to the confirmed ward and how many are discharged back to the community (owing to bed shortage or false-negative diagnosis when using the RDT-only strategy). In addition, and again for each testing strategy, we determine how many patients without Ebola who seek care were infected during their stay in the hospital (in the holding and/or confirmed wards).

To evaluate the impact of a testing strategy from a patient perspective, we compare the CFR among people with suspected Ebola seeking care with the CFR among those same patients had they not sought care (community CFR). Assuming that care reduces the CFR, hospitalization improves the outcome for infected patients. However, patients without Ebola within a health-care unit are at risk of nosocomial transmission; hence hospitalization is not necessarily optimal for a patient with unknown Ebola infection status. We use our model to explore these trade-offs and to determine the epidemiological contexts in which the different testing strategies are optimal.

We assume a CFR of 60% among patients infected with Ebola throughout[18]. Patients without Ebola seeking care in health-care units are likely to present with similar symptoms to those with the infection, including high fever, vomiting, diarrhoea and haemorrhaging. These are symptoms that are typically observed among patients with severe Lassa fever, a disease highly prevalent in West Africa. Therefore, we assume a 20% CFR among patients without Ebola, which is comparable with the reported CFR for severe Lassa fever cases admitted to hospital[19]. Furthermore, we assume that all patients admitted to the confirmed ward (including those without Ebola) benefit from hospital care, with r being the relative CFR of a hospitalized patient with Ebola ($r = 1$ representing no benefit of hospitalization). All patients sent to the confirmed ward are assumed to stay in the health-care unit for an average of 7 days.

The second important role of health-care units in Ebola outbreaks is to reduce transmission by isolating cases. To evaluate the impact of testing strategies on transmission at the level of a single health-care unit, we use our health-care-unit-level model to calculate the reproduction number of Ebola-infected patients seeking care, R_{HU} (the average number of secondary infections generated by these patients). This reproduction number reflects the potential impact of health-care units in reducing transmission and depends on the testing strategy used. We assess the epidemiological contexts for which different testing strategies are optimal.

In examining both the patient and transmission perspectives, we explore likely scenarios at four different stages of the epidemic: in the early growth phase, at the peak, after the peak, and in the tail of the outbreak (reflecting the situation around June 2014, November 2014, January 2015 and May 2015, respectively). Over the course of the epidemic, we assume that the incidence of true cases increases, peaks and then declines; the number of patients without Ebola seeking care initially increases (for example, due to rising awareness as case numbers increase), and remains high as the epidemic wanes (meaning the prevalence of Ebola infection among suspected cases seeking care, p, wanes over time); the reproduction number among those not seeking care

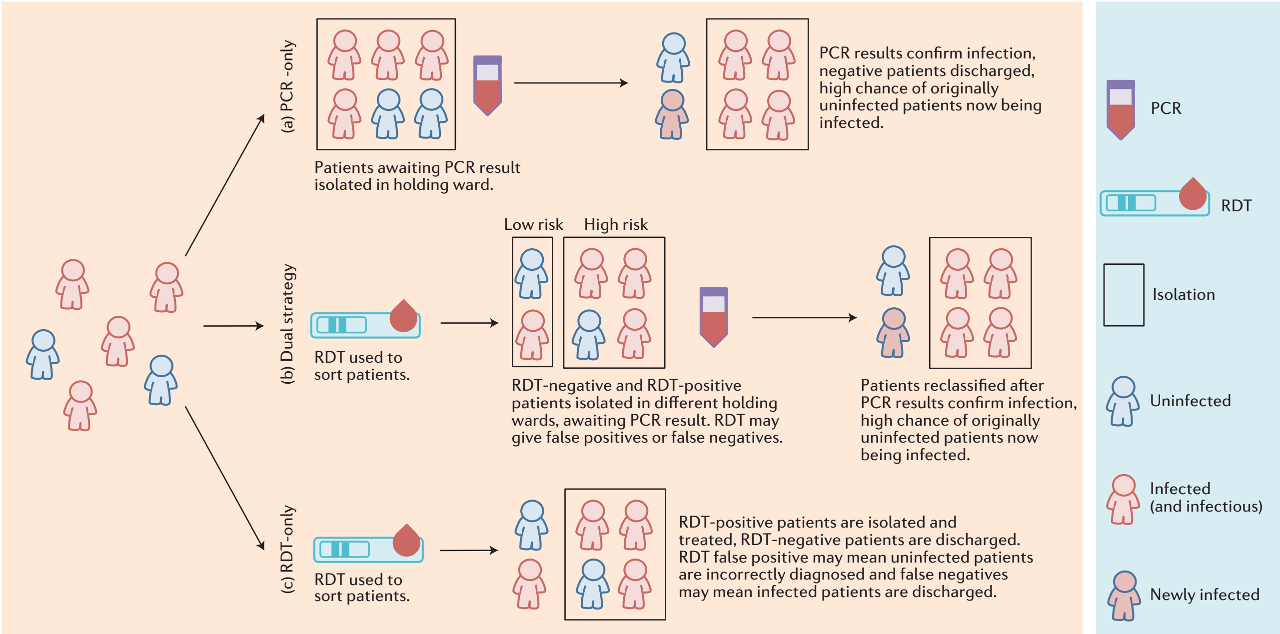


Figure 1 | Three Ebola health-care-unit diagnostic testing and patient triage strategies. Patients seeking care are first admitted to a holding area where they wait to be either admitted to a confirmed ward or discharged back to the community. We examine three diagnostic testing strategies. **a**, Polymerase chain reaction (PCR)-only: patients await their test results in a single holding area. When PCR test results become available, individuals are either sent to a confirmed ward or discharged back to the community. **b**, Dual strategy (rapid diagnostic test (RDT) and PCR): based on initial RDT results, patients seeking care are kept in separate high- or low-risk wards. When PCR test results become available, individuals are either sent to a confirmed ward or discharged back to the community. **c**, RDT-only: the RDT result alone determines who is sent to a confirmed ward or discharged back to the community. Patients are either infected (red), uninfected (blue) or exposed within the holding area (infected, but not yet infectious, blue outline and red centre). The RDT is assumed to have a lower sensitivity and specificity than PCR, therefore RDT may give incorrect classifications (false positives and false negatives), which can result in nosocomial infections when non-Ebola cases are admitted, or a high risk of further transmission in the community if true Ebola cases are erroneously discharged.

(community reproduction number) decreases, reflecting improved community control of transmission (for example, safer funeral practices); bed capacity increases and then plateaus once the epidemic starts to decline (see Table 2 for parameter values).

We implemented the health-care-unit-level model as an Excel spreadsheet and Java program (see Supplementary Information). This freely available software allows the reader to explore the impact of different model assumptions and parameter values on model outcomes.

Potential impact of RDTs at the population level

To evaluate the potential impact of RDT use on the overall trajectory of an Ebola epidemic at the population level (compared with the impact at an individual- or health-care-unit level) we developed a susceptible–exposed–infectious–recovered (SEIR)-type transmission model, which was extended to include a highly infectious 'near death' stage reflecting increased transmission around the time of death and at funerals (see Supplementary Information 3.1.1). The model incorporates observed delays between key epidemiological events and is parameterized to reproduce the basic reproduction number, R_0, observed during the Sierra Leone epidemic[18,20] (model details are given in Supplementary Information 3).

We investigate the same diagnostic strategies as in the health-care-unit-level model (Fig. 1), but allow for full disease progression within each ward and in the community, with the flows between community and hospital wards being determined by the testing delays and test characteristics. As in the health-care-unit-level model, bed capacity and utilization are tracked such that once all beds are occupied, patients seeking care cannot be admitted and therefore remain in the community.

The model incorporates observed changes in the bed capacity available during the epidemic, making use of data on the number of ETU beds opened over time from the start of the outbreak to the end of May 2015. We include a constant number of 60 beds throughout to represent informal holding units and other health-care facilities that were available before the start of the epidemic and were used for Ebola patients before the scale-up of ETU beds. We vary the mean onset to hospitalization delay each month to match reported performance indicators[21]. At the start of the epidemic, the relative infectiousness of a non-hospitalized person with Ebola during the late stage of infection close to, and shortly after, death is assumed to be 16-fold higher than infectiousness earlier in disease progression, but this factor is allowed to decrease over time. Higher late-stage infectiousness allows the model to represent the enhanced risk of transmission to those caring for dying patients at home, preparing the corpse of a family member or undertaking funeral rites. We chose this value to yield a similar overall probability of onward transmission from fatalities as that of survivors despite the shorter infectious period of fatal cases. Thus, around half of community transmission from a non-hospitalized fatal case occurs around the time of death. We assume safe handling of corpses and burials for all deaths in confirmed wards of health-care units.

Assuming no use of RDTs, transmission parameters are calibrated to match the time series of observed incidence (confirmed and probable cases) and bed occupancy in Sierra Leone. The parameters calibrated include R_0, the initial number of infectious cases present at the start of the simulation, the daily rate of hospitalizations of people without Ebola, the probability of care seeking for infected patients and the decrease in death-associated excess transmissibility in the community. These last two parameters were allowed to change at three time points: 1 October 2014, 1 November 2014 and 1 January 2015. The rate of people without Ebola seeking care is assumed to be proportional to the cumulative number of Ebola deaths up to that point.

For the baseline scenario in both health-care-unit-level and population-level models, we assume 100% sensitivity and specificity of PCR testing, and 92% sensitivity and 85% specificity for the RDT, matching the published performance of the ReEBOV test[15]. The average delays in obtaining RDT and PCR results are assumed to be 1 hour and 2 days, respectively.

Table 2 | Illustration of health-care-unit model predictions for the CFR among patients seeking care (relative to the community CFR) and for the reproduction number of true Ebola patients seeking care for three testing strategies for levels of bed demand and true Ebola infection prevalence appropriate for four different stages of the current epidemic, as informed by the calibrated parameters of the population-level transmission model.

Test name	Early			During the peak			Decreasing			Going to zero		
	PCR-only	Dual	RDT-only	PCR-only	Dual	RDT-only	PCR-only	Dual	RDT-only	PCR-only	Dual	RDT-only
ν (daily rate of arrival of patients)‡	5			60			60			60		
p (%, prevalence) ‡	90			70			50			10		
Community reproduction number ‡	1.7			1.7			0.85			0.85		
CFR among patients seeking care*	0.73	0.72§	0.74	0.79	0.76§	0.78	0.85	0.81§	0.84	0.98	0.96§	1.00
R_{HU}*	0.48	0.47§	0.56	0.53	0.49§	0.59	0.36	0.28§	0.37	0.49	0.4§	0.75
Bed demand	33		30	330		290	270		226	150		97
Bed capacity†, ‡	30			200			350			350		
CFR among patients seeking care†	0.75	0.74	0.74§	0.87	0.86	0.85§	0.85	0.81§	0.84	0.98	0.96§	1.00
R_{HU}†	0.57	0.56§	0.56	0.99	0.96	0.94§	0.36	0.28§	0.37	0.49	0.4§	0.75

*Results assuming bed capacity (total number of beds in health-care unit) exceeds demand. †Results assuming health-care-unit bed numbers are limited. For the CFR results, we assume hospitalization decreases the CFR of patients admitted to the confirmed ward by a factor of 0.7. ‡Assumed model parameters. §The optimal strategies. Further model parameters are fixed at their baseline values (Supplementary Tables 4 and 6 show equivalent results with lower and higher nosocomial transmission rate). CFR, case fatality ratio; PCR, polymerase chain reaction; RDT, rapid diagnostic test.

We also examine the impact of reducing the delay in obtaining PCR test results from 2 days to 1 day, the minimum delay realistically achievable when laboratory facilities are located close to health-care units. Although current RDTs have limited sensitivity and specificity, the next generation of diagnostic tests may have substantially improved characteristics. We therefore also investigate the impact of using near-perfect RDTs, assuming sensitivity and specificity of 99% each.

Given the challenges faced during the epidemic such as sample collection, storage, transport and laboratory processing, it is unlikely that the high nominal sensitivity and specificity of PCR testing was achieved. Moreover, considerable variation in the sensitivity and specificity of ReEBOV RDT has been reported in the literature[14–16]. As a sensitivity analysis, we repeat the simulations using arguably more-realistic assumptions, regarding test sensitivity and specificity. We assume PCR test sensitivity and specificity of 85% and 95%, respectively, and RDT sensitivity and specificity of 82% and 80%, respectively, and recalibrate the remaining model parameters. The larger relative decreases in sensitivity and specificity of the PCR test compared with the RDT reflect the greater logistical constraints associated with the former.

RESULTS

After calibrating the population-level transmission model, we estimate that 37% of people with Ebola sought health care at the start of the epidemic, increasing to 41%, 62% and 73% at the beginning of October 2014, November 2014 and January 2015, respectively. The relative risk of transmission associated with death is high up to the peak of the epidemic (16 until 30 September, and 15 until 31 October), and low after the epidemic peak (1.2 between 1 November and 31 December, and 1.1 from January onwards). The rate of patients without Ebola seeking care equates to around 120 patients per day at the end of the epidemic. The calibrated value of R_0 is 1.7 (see Supplementary Table 11 and Supplementary Information 3.4 for sensitivity analyses with regards to the calibrated model parameters).

Potential impact of RDTs in a health-care unit

From a patient perspective, hospitalization is not necessarily optimal and a trade-off exists between the risk of nosocomial infection and the benefits of treatment. At different stages of the epidemic and for the three testing strategies, Figure 2 shows the average CFR per person with suspected Ebola who seeks care as a function of the CFR reduction in the confirmed ward and of the ratio of bed capacity to patients seeking care per day. For hospitalization to be beneficial, the CFR among patients seeking care must be sufficiently reduced by treatment to compensate for the risk of nosocomial transmission to patients without Ebola (Fig. 2). When most patients seeking care are true Ebola cases, a small benefit of treatment is sufficient to reduce the average CFR per patient: at the peak (p = 0.7) using the dual strategy, CFR with treatment must fall below 97% of its community value (r = 0.97) to decrease the average CFR among patients seeking care (Table 2 and Fig. 2e). When most patients seeking care do not have Ebola, a higher benefit of treatment is required to reduce the average CFR per patient: at the tail of the outbreak (p = 0.1) using dual strategy, CFR with treatment must fall below 86% of its community value (r = 0.86) to decrease the average CFR among patients seeking care (Fig. 2k).
We first consider the impact of testing strategies when the demand for beds does not exceed capacity (Fig. 2). When hospitalization has little impact on CFR, the dual strategy is preferred as it reduces nosocomial transmission. However, RDT-only may become preferable (for example, at the tail of an epidemic; Fig. 2j–l) if the health-care-unit CFR is sufficiently lower than the community CFR such that the higher level of nosocomial transmission in the holding wards seen for the dual strategy outweighs the impact of increased nosocomial transmission in the confirmed wards seen for RDT-only, given that the patients in the latter will benefit from care.

When demand for beds exceeds capacity (Fig. 2), the same reasoning applies and RDT-only is increasingly favoured as it allows a larger proportion of cases to be admitted compared with the other strategies.

For realistic parameter values, we estimate that the RDT-only strategy would be marginally optimal during the growing phase of an epidemic, whereas the dual-testing strategy would be marginally optimal in the later stage (Fig. 2, Table 2, Supplementary Table 3). For instance, during the peak of the epidemic, we assume 60 patients seek care daily and 200 beds are available (resulting in 3.3 beds per patient seeking care), a level of capacity that results in 40% of patients being turned away from the health-care units for the PCR-only (and dual) strategy, compared with 32% patients turned away using RDT-only. Furthermore, if we assume a 30% decrease in CFR for those patients sent to the confirmed ward (r = 0.7), the CFR among people with suspected Ebola seeking care (relative to community CFR) would be reduced by 15% under the RDT-only strategy and by 14% under the dual strategy (Table 2, Fig. 2, Supplementary Table 4 for further sensitivity analyses).

Figure 3 shows the impact of introducing RDTs on the transmission by a person with Ebola who seeks care for a scenario that is comparable with that in many affected areas at the peak of the recent West African epidemic (see Supplementary Figs 1–4 for results at other stages of the epidemic). We evaluate the dual testing (Fig. 3a) and RDT-only (Fig. 3b) strategies relative to the impact of PCR-only testing.

When bed capacity exceeds demand, the results illustrate some key conclusions about the impact of RDTs on Ebola transmission: transmission is always lowest for the dual strategy, with a reduction in the reproduction number of up to 14% compared with PCR-only (Fig. 3a, Table 2); depending on the RDT's sensitivity and specificity, the RDT-only strategy may result in lower

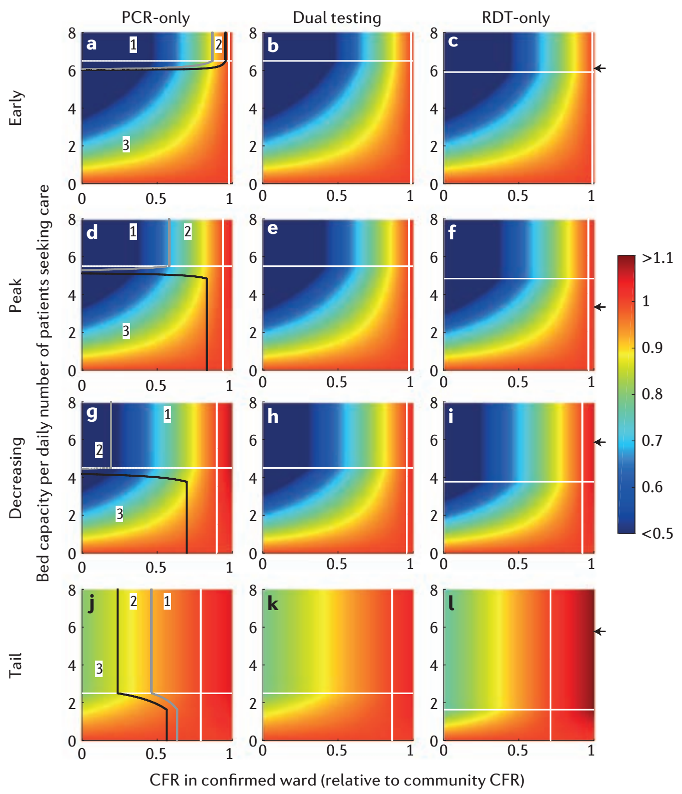


Figure 2 | Case fatality ratio (CFR) of patients seeking care divided by the CFR if those same Ebola and non-Ebola cases had remained in the community. Each row represents a particular stage in the epidemic: from left to right: early (**a,b, c**), during the peak (**d,e,f**), shortly after the peak (**g,h,i**) and once the epidemic is tailing off (**j,k,l**). Each column reflects a testing strategy, namely polymerase chain reaction (PCR)-only (**a,d,g,j**), dual strategy (**b,e,h,k**) and rapid diagnostic test (RDT)-only (**c,f,i,l**). White horizontal lines show the threshold bed capacity below which demand cannot be met for PCR-only (same threshold as dual strategy) and RDT-only. Solid grey and black lines (left panels, **a,d,g,j**) indicate, respectively, where the outcomes of PCR-only and RDT-only are equivalent, and where the outcomes of dual (RDT and PCR) testing and RDT-only are equivalent. Those lines delimit parameter space where (1) dual strategy is best followed by PCR-only and then RDT-only, (2) dual strategy is best followed by RDT-only and then PCR-only and (3) RDT-only is best followed by dual strategy and then PCR-only. On the left of the white solid vertical line (specific for the testing strategy), the benefit of care is sufficient to decrease the average CFR among patients seeking care (unaware of their disease status, and assuming hospital infection control has not improved over the course of the epidemic). The black arrows on the right *y*-axis of the RDT-only plots indicate the likely availability of beds at the corresponding stage of the epidemic (Table 2); however, this is likely to have varied between different health-care units.

transmission than the PCR-only strategy (Fig. 3b, Table 2). Low RDT sensitivity worsens the predicted outcome of the RDT-only strategy much more than the outcome of the dual strategy (Fig. 3a,b).

When bed capacity cannot meet demand, the RDT-only strategy may give lower transmission than both the PCR-only and the dual strategies (Figs 3c, d, Table 2). Relying on RDTs alone is optimal during the peak of the epidemic owing to the strategy's better use of bed capacity (Table 2, Fig. 3c,d, Supplementary Information 2.3, Supplementary Figs 2–5). The benefit of the RDT-alone strategy increases as the level of infection control in the health-care unit decreases (for example see Supplementary Table 6 for results with the rate of nosocomial transmission set at 0.1 and 0.2 per day). Although the reduction in the reproduction number may seem modest during the growing phase of the epidemic (under 11%; Table 2, Supplementary Fig. 1), a reduction of this size can have a considerable impact on cumulative case numbers.

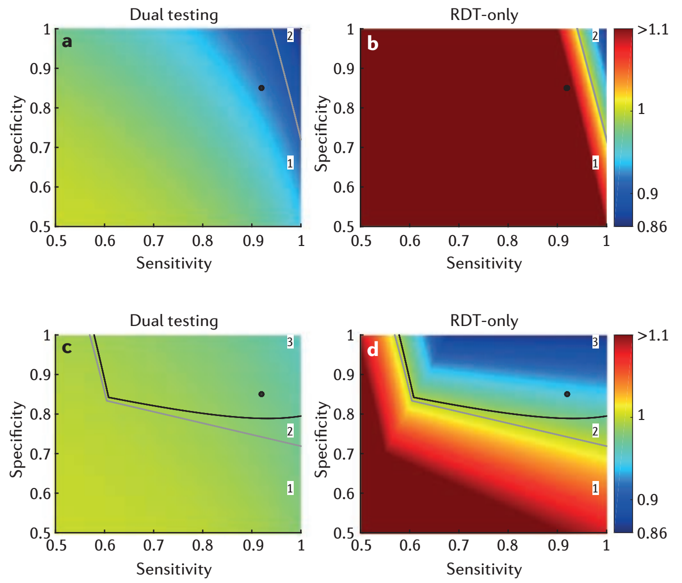


Figure 3 | Relative reproduction number of patients with Ebola who are seeking care for dual strategy and rapid diagnostic test (RDT)-only, compared to polymerase chain reaction (PCR)-only, at the peak of the epidemic (see Table 2 for parameters). For **a,b** bed capacity is unlimited, whereas for **c,d** the health-care unit has 200 beds (Table 2). The outcome using a dual (RDT and PCR) strategy is shown in **a,c**, whereas **b,d** present the outcome using RDT-only. For the specified parameters and when bed capacity is unlimited, the reproduction number for the PCR-only strategy is 0.53 (0.99 when bed capacity is limited to 200). The PCR-only outcome is independent of the RDT's sensitivity and specificity. Solid grey and black lines indicate, respectively, where the outcomes of PCR-only and RDT-only are equivalent, and where the outcomes of dual (RDT and PCR) testing and RDT-only are equivalent. Those lines delimit parameter space where (1) dual strategy is best followed by PCR-only and then RDT-only, (2) dual strategy is best followed by RDT-only and then PCR-only and (3) RDT-only is best followed by dual strategy and then PCR-only. The black circle indicates the World Health Organization reported sensitivity and specificity of the ReEBOV RDT (92% and 85%, respectively)[15].

Potential impact of RDTs at the population level

We achieve a good match between observed incidence in Sierra Leone and the population-level transmission model after calibration, assuming PCR-only testing (Fig. 4a). We then use the model to evaluate the potential impact of RDTs (Fig. 4, Table 3). We estimate that adopting an RDT-only strategy would have reduced the size of the epidemic by 6%, despite imperfect sensitivity leading some patients with Ebola to be discharged back to the community and continuing to transmit infection there, particularly in the tail of the outbreak. The number of patients with Ebola discharged from hospital is similar under both the PCR-only and RDT-only strategies; for the PCR-only strategy, these are nosocomial infections, whereas for the RDT-only strategy they are predominantly cases with false-negative test results. When demand for beds exceeds capacity, the quick turnaround time of the RDT considerably reduces the need for hospital beds, and consequently fewer patients are turned away from hospital (Fig. 4b and Supplementary Fig. 8). Thus, RDT-only might be the preferred testing strategy, particularly when bed demand is exceeded.

However, with the limited sensitivity and specificity of existing RDTs, adopting a dual strategy would have been more effective than relying on either PCR- or RDT-testing alone. We estimate that such a dual strategy could have reduced the size of the epidemic by 32% compared with the PCR-only strategy (Fig. 4, Table 3). Note that the precise reduction in case numbers achieved crucially depends on the risk of nosocomial transmission. The segregation of cases into high- and low-risk wards sufficiently reduces nosocomial transmission to substantially reduce the number of newly infected cases being discharged back to the community. This more efficient disruption of transmission within both health-care-unit and community settings leads to lower incidence early in the epidemic, and consequently a lower bed demand

Table 3 | Summary statistics of the different testing scenarios considered. The upper rows assume test performance as in the baseline scenario, whereas the lower rows assume real-world use leads to a reduction in test performance as described in the methods.

Scenario		Total number of cases (% relative to PCR-only)	Peak weekly incidence	Total number of infected patients discharged	Total number of Ebola cases turned away from health-care units due to lack of beds
Perfect PCR and nominal RDT characteristics	PCR-only	11,600 (100)	620	780	490
	RDT-only	10,800 (94)	580	960	60
	Dual testing	7,900 (68)	460	280	40
	Fast PCR	8,100 (70)	480	320	0
	Near perfect RDT-only	6,700 (58)	390	70	0
'Real world' PCR and RDT characteristics	PCR-only	11,500 (100)	630	1,820	710
	RDT-only	10,700 (93)	600	1,760	260
	Dual testing	9,100 (79)	520	1,270	290
	Fast PCR	8,800 (77)	540	1,260	70

PCR, polymerase chain reaction; RDT, rapid diagnostic test

throughout, reducing the time periods when bed demand exceeds capacity (Fig. 4b, Supplementary Fig. 9).

Reducing the average delay in obtaining PCR test results from 2 days to 1 day could have reduced overall case numbers by 30% without recourse to RDTs (Fig. 4, Table 3). The impact of faster PCR testing is twofold: a reduction in the average stay in health-care units reduces demand for beds in the holding wards, and reduces nosocomial transmission.

If near-perfect RDTs were available (with both sensitivity and specificity at 99%), we estimate that the epidemic in Sierra Leone could have been reduced by 42% — substantially better than even 1-day-turnaround PCR testing could achieve. In this scenario, the duration of health-care-unit stay for people who are suspected of having, but are not infected with, Ebola is minimal, implying a considerably lower demand for beds (Fig. 4b). The near-perfect RDT generates a very small number of misdiagnoses, therefore reducing both the numbers of false-negative discharges and false-positive admissions to the confirmed ward — saving available beds for patients truly infected with Ebola (Supplementary Fig. 9).

Overall, model predictions of the epidemic size are highly sensitive to the assumed values of RDT sensitivity and specificity for the RDT-only strategy, ranging from a 42% reduction for a near perfect test to nearly a twofold increase if sensitivity and specificity are very poor. However, model predictions for the dual strategy are much less sensitive to the RDT characteristics, with the reduction in epidemic size only varying in the range 27% to 41% (see Supplementary Information 3.4).

Sensitivity analyses that assume lower PCR and RDT sensitivity and specificity (owing to challenges in sample-taking and transport) give qualitatively similar results (Table 3).

DISCUSSION

Our results support the WHO advice on the use of RDTs, showing that RDTs could improve the control of Ebola epidemics, particularly in contexts in which laboratory or bed capacity is limited compared with demand, and infection control between patients in health-care units is imperfect. Implementing a dual testing strategy (whereby RDTs are used for early triage into high- and low-risk holding areas, followed by confirmation by PCR before admission to confirmed wards or patients are discharged) has the potential to decrease nosocomial transmission, leading to a sizeable reduction in the final epidemic size.

Crucial to the assessment of the benefits of RDTs in Ebola control is the ever-present risk of nosocomial transmission. Even when health-care-unit bed capacity can meet demand, a trade-off exists between using a slow PCR test and relying solely on RDTs. The longer wait for PCR test results increases the risk that suspected patients who are not actually infected with Ebola will become infected in holding areas. However, relying on RDTs alone has two potential negative consequences: false positives due to imperfect specificity, which leads to increased nosocomial transmission as non-Ebola patients are sent to the confirmed ward; and false negatives due to imperfect sensitivity, leading to discharge of infected individuals back to the community. When most people with suspected Ebola seeking care are infected (as might be expected in the growth phase of an epidemic), high sensitivity is the more important of these two factors to reduce the risks of discharging false negatives. When most people with suspected Ebola are not infected with Ebola (as expected in the tail of an epidemic) high specificity is needed to minimize the number of false positives. As nosocomial transmission risks increase (for example, due to high demand on services, overcrowding or limitations of infection control), so do the benefits of adopting RDTs in combination with PCR testing.

The prevalence of Ebola infection in patients seeking care and the risk of nosocomial transmission crucially determine the optimal testing strategy. Intermediate values of the underlying prevalence of Ebola infection among those seeking care give the highest risk of nosocomial transmission by maximizing contact between patients with and without Ebola.

Although the average bed occupancy per suspect case is the same for the PCR-only and dual strategies, the RDT-only strategy affects bed usage: less time is spent awaiting test results whereas the proportion of patients sent to the confirmed ward is determined by RDT sensitivity and specificity. Here, the combination of high sensitivity and specificity is of course optimal — a poor specificity of RDT may increase overall bed occupancy by causing increased numbers of patients without Ebola to be falsely confirmed with Ebola and admitted to the confirmed wards.

During the current epidemic there were periods in which bed demand exceeded capacity. In such circumstances, our analysis suggests that relying on RDTs alone could substantially improve utilization of bed capacity, despite less than perfect diagnostic sensitivity and specificity. We estimate that using the RDT-only strategy throughout the epidemic would have decreased case numbers by a modest 6% in Sierra Leone. This is because the better bed utilization under this strategy is counterweighed by the higher false-positive rate of current RDTs compared with PCR; this becomes more of an issue in the tail of an epidemic during which we expect the prevalence of true Ebola infection among people with suspected Ebola to decline. However, the RDT-only strategy would be a more attractive option if infection control in health-care units could be increased in the tail of an epidemic, mitigating this negative effect.

The dual strategy using both PCR and RDT was predicted to have the greatest impact on epidemic size, potentially reducing the size of the epidemic in Sierra Leone by a third. A similar reduction could have been achieved with PCR testing alone if results had been consistently reported back to health-care units within 24 hours. However, as the logistical challenges involved in setting up rapid, high-throughput PCR testing close to the point of care are substantial, RDTs are likely to be a useful addition to the diagnostic armoury for combatting future Ebola epidemics.

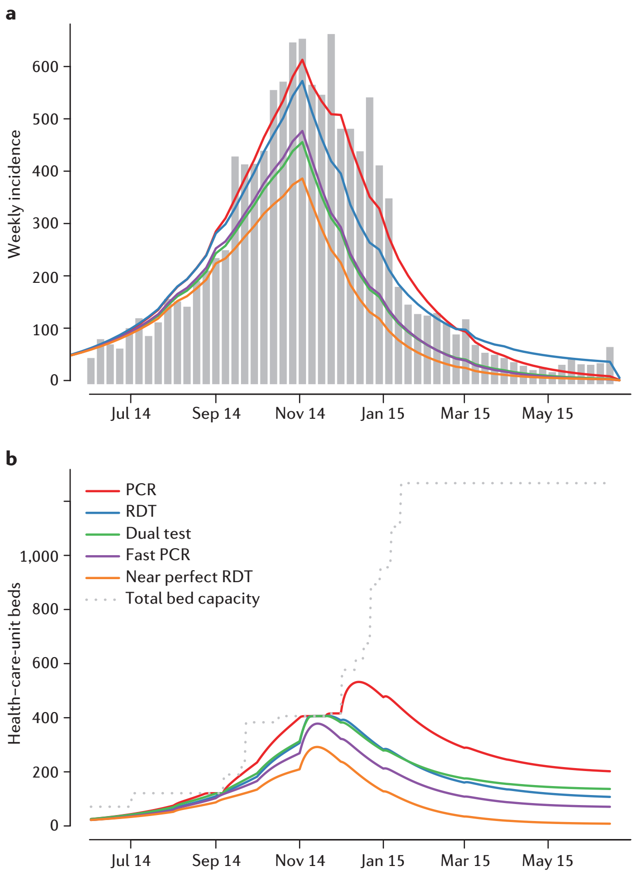


Figure 4 | Ebola outbreak in Sierra Leone. a, Observed (grey bars) and expected (coloured lines) weekly incidence of confirmed and probable Ebola cases during the outbreak in Sierra Leone. The red line presents the expected incidence using the polymerase chain reaction (PCR)-only strategy on which the model was calibrated. Other lines present the estimated incidence under counterfactual scenarios: rapid diagnostic test (RDT)-only (blue), dual strategy (green), PCR-only with faster test results delivered within 1 day (purple) or RDT-only with near-perfect sensitivity and specificity of 99% each (orange). **b,** Bed capacity (grey) and usage (colours) throughout the epidemic in Sierra Leone for the same scenarios as in (**a**).

Our models inevitably simplify the complex processes underlying Ebola transmission dynamics and patient testing and treatment. However, they provide an important initial evaluation of the potential impact of RDT use on outbreak control. Our qualitative conclusions were robust to sensitivity analyses, and although the numerical predictions of reductions in epidemic size should be cautiously interpreted, the benefit of using RDTs in combination with PCR testing remained. In the baseline population level analysis, we assumed perfect sensitivity and specificity of PCR testing, but logistical challenges in field conditions might lead to substantially lower real-world performance[16,22]. However, we found that the predicted qualitative impact of RDTs was robust to possible suboptimal sensitivity and specificity of PCR tests (Table 3).

Furthermore, there is considerable uncertainty in RDT performance, with reported sensitivity and specificity for ReEBOV varying between 78% and 100%, and 85% and 92%, respectively[14–16]. Such differences would have substantial epidemiological and clinical consequences should an RDT-only testing strategy be adopted. This uncertainty also limits our ability to assess the potential impact of RDTs, particularly if used alone. The predicted impact of dual-testing strategies is less sensitive to uncertainty in RDT performance, with all reported values for ReEBOV resulting in a considerable reduction of the epidemic size (between 27% and 41%).

Finally, for the main analysis presented here, we conservatively assumed that RDTs would be less accurate than PCR tests. Although this is true for ReEBOV[14,15] (Supplementary Table 1), some alternative RDTs in development seem to be as accurate as their PCR counterparts, even when field tested (for example, REVAMP[23]; J.-C. Manuguerra, personal communication; Supplementary Table 1). Such tests could have a dramatic impact on the control of future epidemics in which health care or laboratory capacity is limited (Table 3). Therefore as innovation in RDT development continues, RDTs could quickly replace PCR testing, achieving similar accuracy with the benefit of faster results, simpler logistics and safer handling of isolates. This article has evaluated a limited set of diagnostic testing strategies for Ebola and further work should evaluate alternative diagnostic methods tailored to field procedures. In Sierra Leone, the epidemic response was decentralized with each district hosting an Ebola response centre, which investigated local cases based on contact tracing, potential exposure and local knowledge. In such settings, an alternative testing strategy could see RDTs used in the community. This could reduce bed demand, costs, patient-transportation needs and risks of transmission. Furthermore, such community use of RDTs could potentially identify cases that would not have met clinical case definitions, and would send a positive message to the community that the testing is fast and transparent.

In the recent West African Ebola epidemic it has been difficult to obtain information about the changes over time in funeral practices, care-seeking behaviour, bed capacity and population mobility. In addition, the level of case ascertainment is uncertain, but is likely to have been poor at certain times in some areas[11,24,25]. These data gaps may be at least partly filled if priority is placed on collation of the wealth of local knowledge and data that may otherwise become lost as the people who contributed to the response resume other activities.

Updating standard diagnostic and triage procedures for Ebola to include the use of RDTs could offer substantial benefits both from a patient and health-care-unit perspective, and in improving overall control of an epidemic. Until now, Ebola outbreaks were thought to be easily controlled. The usual narratives surrounding the control of Ebola highlight the importance of safe funerals, prompt isolation and effective contact tracing. However, testing strategies can also have a crucial role in minimizing the opportunity for nosocomial transmission and maximizing bed utilization. The ease of use of RDTs, particularly in a resource-poor setting, makes them potentially powerful tools for rapid detection and containment of future outbreaks.

1. World Health Organization. *Laboratory Diagnosis of Ebola Virus Disease* (WHO, 2014).
2. World Health Organization. *Urgently Needed: Rapid, Sensitive, Safe and Simple Ebola Diagnostic Tests* http://www.who.int/mediacentre/news/ebola/18-november-2014-diagnostics/en/ (2014).
3. Altona Diagnostics. *RealStar® Filovirus Screen RT-PCR Kit 1.0.* (Altona, 2014).
4. Obelis, S. A. & Shanghai, Z. J. Bio-Tech Co. Ltd. *LifeRiver Ebola Virus (EBOV) Real Time RT-PCR Kit User Manual* (Bio-Tech, 2012).
5. Saijo, M. *et al.* Laboratory diagnostic systems for Ebola and Marburg hemorrhagic fevers developed with recombinant proteins. *Clinical Vaccine Immunol.* **13,** 444–451 (2006).
6. Towner, J. S. *et al.* Rapid diagnosis of Ebola hemorrhagic fever by reverse transcription-PCR in an outbreak setting and assessment of patient viral load as a predictor of outcome. *J. Virology* **78,** 4330–4341 (2004).
7. Strecker, T. *et al.* Field evaluation of capillary blood samples as a collection specimen for the rapid diagnosis of Ebola virus infection during an outbreak emergency. *Clinical Infect. Dis.* **61,** 669–675 (2015).
8. Chua, A. C., Cunningham, J., Moussy, F., Perkins, M. D. & Formenty, P. The case for improved diagnostic tools to control Ebola virus disease in West Africa and how to get there. *PLoS Negl. Trop. Dis.* **9,** e0003734 (2015).
9. Zachariah, R. & Harries, A. D. The WHO clinical case definition for suspected cases of Ebola virus disease arriving at Ebola holding units: reason to worry? *Lancet Infect. Dis.* **15,** 989–990 (2015).
10. Pathmanathan, I. *et al.* Rapid assessment of Ebola infection prevention and control needs — six districts, Sierra Leone, October 2014. *Morb. Mortal. Wkly Rep.* **63,** 1172–1174 (2014).
11. World Health Organization. *Ebola Situation Report — 26 November 2014* http://apps.who.int/ebola/en/ebola-situation-report/situation-reports/ebola-situation-report-26-november-2014 (WHO, 2014).
12. World Health Organization. *Ebola Situation Report — 21 January 2015* http://apps.who.int/ebola/en/status-outbreak/situation-reports/ebola-situation-report-21-january-2015 (WHO, 2015).
13. World Health Organization. *WHO Emergency Quality Assessment Mechanism for EVD IVDs. Product: RealStar® Filovirus Screen RT-PCR Kit 1.0* (WHO, 2015).
14. Corgenix. *ReEBOV Antigen Rapid Test (Ebolavirus VP40 Antigen Detection) Instructions for Use* (Corgenix, 2015).
15. World Health Organization. *WHO Emergency Use Assessment and Listing for Ebola Virus Disease IVDs. Product: ReEBOV Antigen Rapid Test Kit* (WHO, 2015).
16. Broadhurst, M. J. *et al.* ReEBOV Antigen Rapid Test kit for point-of-care and laboratory-based testing for Ebola virus disease: a field validation study. *Lancet* **386,** 867–874 (2015).

17. Food and Drug Administration. *2014 Ebola Virus Emergency Use Authorizations* http://www.fda.gov/medicaldevices/safety/emergencysituations/ucm161496.htm#ebola (FDA, 2014–2015).
18. WHO Ebola Response Team. West African Ebola epidemic after one year — slowing but not yet under control. *N. Engl. J. Med.* **372,** 584–587 (2015).
19. National Center for Emerging and Zoonotic Infectious Diseases CDC. *Lassa Fever* http://www.cdc.gov/vhf/lassa/pdf/factsheet.pdf (CDC, 2015).
20. WHO Ebola Response Team. Ebola virus disease in West Africa — the first 9 months of the epidemic and forward projections. *N. Engl. J. Med.* **371,** 1481–1495 (2014).
21. World Health Organization. *Ebola Situation Report — 27 May 2015* http://apps.who.int/ebola/en/current-situation/ebola-situation-report-27-may-2015 (WHO, 2015).
22. Bhadelia, N. Rapid diagnostics for Ebola in emergency settings. *Lancet* **386,** 833–835 (2015).
23. Manuguerra, J.-C. Molecular pathogen detection in the field, directly at the point of care. *Proc. 3rd Int. Congress on Targeting Infectious Diseases* (Task Force Infectious Disease, 2015).
24. Meltzer, M. I. *et al.* Estimating the future number of cases in the Ebola epidemic — Liberia and Sierra Leone, 2014-2015. *Morb. Mortal. Wkly. Rep.* **63,** 1–14 (2014).
25. World Health Organization. *Ebola Situation Report — 14 January 2015* http://apps.who.int/ebola/en/status-outbreak/situation-reports/ebola-situation-report-14-january-2015 (2015).
26. Cepheid. *Cepheid Xpert Ebola Assay* (Cepheid, 2015).
27. World Health Organization. *WHO Emergency Use Assessment and Listing for EVD IVDs PUBLIC REPORT Product: Xpert® Ebola Assay* (WHO, 2015).
28. World Health Organization. *WHO Emergency Use Assessment and Listing Procedure for EVD IVDs. PUBLIC REPORT. Product: Liferiver™ — Ebola Virus (EBOV) Real Time RT-PCR Kit* (WHO, 2015).

SUPPLEMENTARY MATERIAL

Is linked to the online version of this paper at: http://dx.doi.org/10.1038/nature16041

ACKNOWLEDGEMENTS

We thank J.-C. Manuguerra for sharing preliminary results on the effectiveness of the RE-VAMP RDT and L. Simonsen for early discussions. We furthermore acknowledge research funding from the Medical Research Council, the Bill and Melinda Gates Foundation, the MIDAS network of the National Institute of General Medical Sciences (National Institutes of Health), the Health Protection Research Units of the National Institute for Health Research, the European Union Seventh Framework Programme [FP7/2007–2013] under Grant Agreement no 278433-PREDEMICS, and the Wellcome Trust.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests. Financial support for this publication has been provided by the Bill & Melinda Gates Foundation.

ADDITIONAL INFORMATION